



QY 87 NLPLRFRNVOERSAGATANLPLSRGRNMEVSLVRRV--PNLPORFGRTTTAKSVCRML 144  
Db 266 MVAVKM---LKEASADMOADF-----OREAALMAEFDNPVIVKLLGVCVAGKPMCLLF 316  
QY 145 SDCQSGMHSPCANDLFYSMT 165  
Db 317 EYMGDL-----NEFLRSM 332

## RESULT 2

PCT-US95-08493-19  
; Sequence 19, Application PC/TUS9508493  
; GENERAL INFORMATION:  
; APPLICANT: Wood, Clive  
; APPLICANT: Caruso, Anthony  
; TITLE OF INVENTION: Novel mlk Receptor Tyrosine Kinases  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEGAL AFFAIRS  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/08493  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brown, Scott A  
; REGISTRATION NUMBER: 32,724  
; REFERENCE/DOCKET NUMBER: G15234A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8224  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 860 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US95-08493-19

Query Match 7.5%; Score 76.5; DB 5; Length 860;  
Best Local Similarity 21.4%; Pred. No. 3.4;  
Matches 43; Conservative 32; Mismatches 63; Indels 63; Gaps 13;

QY 1 MEISS-KLFIILLTLATSLTSLTSPNIFCA-----LYCRRRKKNKKKREAVTLTLPSELLDLRLHP 541  
Db 489 ISIVSSFALLTAT-----LYCRRRKKNKKKREAVTLTLPSELLDLRLHP 541  
QY 38 KENYDKYS---EPR---GYPKGERSLNFEELKDWGP---KNVLMSTPAVKNMPS-FA 86  
Db 542 NPMYQRMPLLLNPKLLSLEYPRN---NIEYVRDIGEGAFGRVFOARAPGL--LPYEPFT 595  
QY 87 NLPLRFRNVOERSAGATANLPLSRGRNMEVSLVRRV--PNLPORFGRTTTAKSVCRML 144  
Db 596 MVAVKM---LKEASADMOADF-----OREAALMAEFDNPVIVKLLGVCVAGKPMCLLF 646  
QY 145 SDCQSGMHSPCANDLFYSMT 165  
Db 647 EYMGDL-----NEFLRSM 662

## RESULT 3

PCT-US95-08493-21  
; Sequence 21, Application PC/TUS9508493

; GENERAL INFORMATION:  
; APPLICANT: Wood, Clive  
; APPLICANT: Caruso, Anthony  
; TITLE OF INVENTION: Novel mlk Receptor Tyrosine Kinases  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEGAL AFFAIRS  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/08493  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brown, Scott A  
; REGISTRATION NUMBER: 32,724  
; REFERENCE/DOCKET NUMBER: G15234A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 498-8224  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 868 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US95-08493-21

Query Match 7.5%; Score 76.5; DB 5; Length 868;  
Best Local Similarity 21.4%; Pred. No. 3.5;  
Matches 43; Conservative 32; Mismatches 63; Indels 63; Gaps 13;

QY 1 MEISS-KLFIILLTLATSLTSLTSPNIFCA-----LYCRRRKKNKKKREAVTLTLPSELLDLRLHP 549  
Db 497 ISIVSSFALLTAT-----LYCRRRKKNKKKREAVTLTLPSELLDLRLHP 549  
QY 38 KENYDKYS---EPR---GYPKGERSLNFEELKDWGP---KNVLMSTPAVKNMPS-FA 86  
Db 550 NPMYQRMPLLLNPKLLSLEYPRN---NIEYVRDIGEGAFGRVFOARAPGL--LPYEPFT 603  
QY 87 NLPLRFRNVOERSAGATANLPLSRGRNMEVSLVRRV--PNLPORFGRTTTAKSVCRML 144  
Db 604 MVAVKM---LKEASADMOADF-----OREAALMAEFDNPVIVKLLGVCVAGKPMCLLF 654  
QY 145 SDCQSGMHSPCANDLFYSMT 165  
Db 655 EYMGDL-----NEFLRSM 670

## RESULT 4

US-08-374-834-16  
; Sequence 16, Application US/08374834  
; Patent No. 5656473  
; GENERAL INFORMATION:  
; APPLICANT: Valenzuela, et al.  
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTOR  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
; STREET: 777 Old Saw Mill River Road  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,834
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095,658
; FILING DATE: 21-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J.
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 190A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 345-7400
; TELEFAX: (914) 345-7721
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 869 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-374-834-16

```

```

Query Match          7.4% Score 75.5; DB 1; Length 869;
Best Local Similarity 20.1%; Pred. No. 4.6;
Matches 44; Conservative 37; Mismatches 69; Indels 69; Gaps 14;

QY 1 MEIIS-KLFILLTLATSSLLTSNIFCA-----LYCCRRKQKKNKKRESAAVTLTLPSELLDLRLHP 550
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 498 ISIMSSFAIFVLLTIT-----LYCCRRKQKKNKKRESAAVTLTLPSELLDLRLHP 550

QY 38 KENVDKYS---EPR---GYPKGRSLNFEELKDWGP---KNVTKMSTPAVNKMPHS-FA 86
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 551 NPMYQRMPLLNPKLLSLEYPRN---NIEVVRDIGEGAFGRVFOARAGL--LPYEPFT 604

QY 87 NLPLRFGRNVOEERSAGATANLPLRSGRNNEVSLVRRV--PNLPQRFGRTRTTAKSVCRML 144
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 605 MVAVKM---LKEASADMQADF-----QREAAALMAEFDPNPVILKLGVCAGVKPKMCLLF 655

QY 145 SDLCQCG-----SMHSPCANDLFYSMTQHQEIQNP 174
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 656 EYNAYGDLNEFLRSMSPHTVCS--LSHSDLSMRAQVSSP 692

```

```

RESULT 5
US-08-644-271-29
; Sequence 29, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; TITLE OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644,271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USN 60/008,657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J.
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 869 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-644-271-29

```

```

Query Match          7.4% Score 75.5; DB 2; Length 869;
Best Local Similarity 20.1%; Pred. No. 4.6;
Matches 44; Conservative 37; Mismatches 69; Indels 69; Gaps 14;

QY 1 MEIIS-KLFILLTLATSSLLTSNIFCA-----LYCCRRKQKKNKKRESAAVTLTLPSELLDLRLHP 550
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 498 ISIMSSFAIFVLLTIT-----LYCCRRKQKKNKKRESAAVTLTLPSELLDLRLHP 550

QY 38 KENVDKYS---EPR---GYPKGRSLNFEELKDWGP---KNVTKMSTPAVNKMPHS-FA 86
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 551 NPMYQRMPLLNPKLLSLEYPRN---NIEVVRDIGEGAFGRVFOARAGL--LPYEPFT 604

QY 87 NLPLRFGRNVOEERSAGATANLPLRSGRNNEVSLVRRV--PNLPQRFGRTRTTAKSVCRML 144
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 605 MVAVKM---LKEASADMQADF-----QREAAALMAEFDPNPVILKLGVCAGVKPKMCLLF 655

QY 145 SDLCQCG-----SMHSPCANDLFYSMTQHQEIQNP 174
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 656 EYNAYGDLNEFLRSMSPHTVCS--LSHSDLSMRAQVSSP 692

```

```

RESULT 6
US-09-077-955-33
; Sequence 33, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-33

```

```

Query Match          7.4% Score 75.5; DB 4; Length 869;
Best Local Similarity 20.1%; Pred. No. 4.6;
Matches 44; Conservative 37; Mismatches 69; Indels 69; Gaps 14;

QY 1 MEIIS-KLFILLTLATSSLLTSNIFCA-----LYCCRRKQKKNKKRESAAVTLTLPSELLDLRLHP 550
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 498 ISIMSSFAIFVLLTIT-----LYCCRRKQKKNKKRESAAVTLTLPSELLDLRLHP 550

```

QY 38 KENTDKYS---EPR---GYPKGRSLNFEELKDWGP---KNVIMKSTPAVNKMPS-PA 86  
DB 551 NPMYFORMPLLNPKLLSLEYPN---NIEYVRDIGEAGRGVFOARAPGL--LPYEPFT 604  
QY 87 NLPLRFGRNVQERSAGATANLPLRSRGNVEVSLVRV--PNLPQRFGRTRTTAKSVCRML 144  
DB 605 MVAVKM---LKEASADMQADF-----QREAAALMAEFDPNPVIVKLLGVCVAGKPMCLLF 655  
QY 145 SLDLCQ-----SMHSPCANDLFYSMTQCHOEIQNP 174  
DB 656 EYMAYGDLNEFLRSMSPHTVCS--LSHSDLSMRAQVSSP 692

## RESULT 7

US-09-594-669-16  
; Sequence 16, Application US/09594669  
; Patent No. 6331424  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; TITLE OF INVENTION: motor proteins and methods for  
; FILE REFERENCE: 1042  
; CURRENT APPLICATION NUMBER: US/09/594,669  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 09/295,612  
; PRIOR FILING DATE: 1999-04-20  
; PRIOR APPLICATION NUMBER: US 09/314,464  
; PRIOR FILING DATE: 1999-05-18  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 725  
; TYPE: PRT  
; ORGANISM: Human  
US-09-594-669-16

Query Match 7.4%; Score 75; DB 4; Length 725;  
Best Local Similarity 21.0%; Pred. No. 4;  
Matches 49; Conservative 43; Mismatches 77; Indels 64; Gaps 13:  
QY 14 LATSSLLTSNIP--CADELVMN--LHSKE-----NYDK-----YSEPRGYPKGRSLNF 59  
DB 1 MAWDSLQARLPGLAIKIQRSNGLIHSANVTNLEKSCVSENAE--GGATKG--KEIDF 58  
QY 60 EELKDWGPKNVIMKSTPAVNKMP--HSFANLPLRFGRNVQERSAGATANLPL-----R 111  
DB 59 DQVAAINPE-----LLOLLPLHPKDNLPLOENVTTOKOKRRSVNSKIPAPKESLSR 110  
QY 112 SGRNMEVSLVR-----RVPNLPQRFGRTRTTAKSVCRMLSDLCQ 149  
DB 111 STRMTSVSELITAOENDMEVELPAAANSRKQFSVPAPTRPSCPAVAEIPLRVSEME 170  
QY 150 GSNH---SPCANDL--FYSMTQCHOEIQNPDOKOSRLLPKKIDDAELKQEK 196  
DB 171 EOVHSIRGSSANPVNSVRKSLVKEVKMKNKREE----KKAQNSEMRMKR 219

## RESULT 8

US-09-749-588-2  
; Sequence 2, Application US/09749588  
; Patent No. 6423521  
; GENERAL INFORMATION:  
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CLO01068  
; CURRENT APPLICATION NUMBER: US/09/749,588  
; CURRENT FILING DATE: 2000-12-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2

; LENGTH: 1170  
; TYPE: PRT  
; ORGANISM: Human  
US-09-749-588-2

Query Match 7.3%; Score 74.5; DB 4; Length 1170;  
Best Local Similarity 22.9%; Pred. No. 9.4;  
Matches 49; Conservative 33; Mismatches 53; Indels 79; Gaps 12;

QY 31 VMSNLHSENKYSEPRGYP-----KGRSLNFEELK---DWGPKNVIMKSTPAVNKMP 82  
DB 237 ILSRL--SSENADEYNFVRSEYECFQHKNHHTCLVFEMLEQNLYDLKON----- 282  
QY 83 HSFANLPLRFGRNVQERSAGATANLPLRS-----GRNMEVSLVRVNPVLPOR-- 130  
DB 283 -XFSPLPLKYIRPILOQ---VATALMKLSGLIHADLAPENIMLVDPVRO---PYRVK 334  
QY 131 ---FGRTT--TAKSVCR-----MLSDLCQSGSMHSPC 156  
DB 335 VIDFGSASHVSKAVCSTYLSQSRYYRAPEIILGLPFCEADMSLGCVIAELFLGMPLYPG 394  
QY 157 ANDLFYSMTQCHOEIQNP--DQKOSRLLPKKIDD 189  
DB 395 ASE--YDQTPPEHELETGIGKSEARKYIFNCLDD 426

## RESULT 9

US-08-933-750C-10  
; Sequence 10, Application US/08933750C  
; Patent No. 5932442  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,750C  
; FILING DATE: September 23, 1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0356 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 348 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

```

; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MUSCNOT01
; CLONE: 118160
; US-08-933-750C-10

Query Match          7.1%   Score 72.5; DB 2; Length 348;
Best Local Similarity 22.2%; Pred. No. 2.7;
Matches 4%; Conservative 25; Mismatches 81; Indels 41; Gaps 8;

QY 26 CADELVMSNLHSHK-ENYDKYSEPRGYPKGERSLNFPEELK-----DWGPKNVIKMST 75
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 59 CSDEVWSLAKVLIKMKRWLLDSGPPKGEGERAKKKEKGLSCSDWKPE----- 111

QY 76 PAVNKMPHSFANLPLRGRNVNQERSAGATANLPLRSGRNMEVSLVRVPNLPORFGRTT 135
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 112 -AGLSPPKKKKREDPKTRRDSYDSKSSSSSPKRPESVRSNSKSKAEB- PKTPSSPLTPT 169

QY 136 TAKSVCRMLSDCOG-SMHSPPC-----ANDLF--YSMTCQ-----HOEION 173
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 170 FASSMCLLAPCYLGTDSVRDCKVEMLSAALKADDDYKDYGVNCDKMAEIEDHIYQELKS 229

QY 174 PDQKQSRRL 182
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 230 TDMKYRNRY 238

RESULT 10
US-09-234-613-10
; Sequence 10, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid

```

```

; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-038-760-3

Query Match          7.1%; Score 72; DB 1; Length 816;
Best Local Similarity 21.5%; Pred. No. 11;
Matches 47; Conservative 23; Mismatches 63; Indels 86; Gaps 8;

QY 17 SSLTSNIF-----CADELVMS-----NLHSENYDKYSEPR 48
Db 362 SKLLNDNFIMSLACALEVMATYSRSTQNLSGDTLSPFWILNVLNKAFDFYKVIE 421
QY 49 GYPKGRSLNPEELK-----DNGPKNVIKMSTPAVNKMPSHSANLPRLR 91
Db 422 SFKAEGNLTREMKNKHLCERHRINESLAWLSDSPLDLIKQSKDRBGTTHLESACPL- 480
QY 92 FGRNVQERSAGATANLP LRSGR---NNEVSLVRVPNLPORFGRTTTAKSVCRM LSDLCL 148
Db 481 -----NLPLQNNHTAADWYLSPVRS----PKKGSTTRVNSTANNATQ-- 519
QY 149 QGSNHSPCANDLFYSMTCHOEIQONPDOKQRRLFLFKKI 187
Db 520 -----ATSAP-----QTQKLKSTSLSLFYKKV 542

RESULT 14
US-08-470-091-3
; Sequence 3, Application US/08470091
; Patent No. 5912236
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong-Ji
; APPLICANT: Benedict, William F.
; TITLE OF INVENTION: Broad-Spectrum Tumor Suppressor Genes, Gene Products and
; TITLE OF INVENTION: Methods for Tumor Suppressor Gene Therapy.
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,091
; FILING DATE: JUN-16-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,760
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 7409-025-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 10:14:34 ; Search time 28 Seconds  
(without alignments)  
923.935 Million cell updates/sec

Title: US-09-831-758a-8

Perfect score: 1018

Sequence: 1 MEIISKFLTLATSSLL.....KQSRRLFFKIDDAELKQEK 196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 613006 seqs, 131990659 residues

Total number of hits satisfying chosen parameters: 613006

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents\_AA\_New.\*

- 1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pcp.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pcp.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pcp.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pcp.\*
- 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pcp.\*
- 6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pcp.\*
- 7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	9.5	931	6	US-10-282-122A-56549
2	84	8.3	629	6	Sequence 56549, A
3	83.5	8.2	1817	6	Sequence 2002, Ap
4	81.5	8.0	1358	6	Sequence 3988, Ap
5	81.5	8.0	1426	6	Sequence 1658, Ap
6	81.5	8.0	1435	6	Sequence 1458, Ap
7	81	8.0	471	5	Sequence 2178, Ap
8	81	8.0	471	5	Sequence 2178, Ap
9	81	8.0	502	5	Sequence 74898, A
10	81	8.0	537	5	Sequence 74898, A
11	81	8.0	537	5	Sequence 74898, A
12	81	8.0	537	5	Sequence 74898, A
13	81	8.0	709	5	Sequence 74897, A
14	81	8.0	713	5	Sequence 74897, A
15	81	8.0	713	5	Sequence 74897, A
16	81	8.0	713	5	Sequence 74897, A
17	81	8.0	772	5	Sequence 74922, A
18	81	8.0	1343	6	Sequence 74921, A
19	81	8.0	1351	6	Sequence 74921, A
20	80	7.9	302	6	Sequence 75965, A
21	80	7.9	556	6	Sequence 75147, A
22	80	7.9	609	6	Sequence 6, Appli
23	80	7.9	609	6	Sequence 4, Appli
24	80	7.9	1305	6	Sequence 3, Appli
25	79.5	7.8	509	6	Sequence 12, Appl
26	79.5	7.8	1275	6	Sequence 77989, A
					Sequence 73404, A
					Sequence 22705, A

27 79 7.8 647 6 US-10-282-122A-71843  
28 78.5 7.7 588 6 US-10-264-213-167  
29 78.5 7.7 660 6 US-10-264-213-138  
30 78.5 7.7 662 6 US-10-264-213-332  
31 78.5 7.7 709 6 US-10-264-213-155  
32 77 7.6 270 6 US-10-282-122A-60255  
33 77 7.6 1131 6 US-10-366-683-31247  
34 76.5 7.5 1357 6 US-10-369-493-5432  
35 76 7.5 899 6 US-10-369-493-4010  
36 75.5 7.4 460 5 US-09-724-676-82376  
37 75.5 7.4 460 5 US-09-724-676-82376  
38 75 7.4 725 5 US-09-849-602-20  
39 74.5 7.3 280 1 PCT-US03-03161-2  
40 74.5 7.3 280 6 US-10-350-385-2  
41 74.5 7.3 490 1 PCT-US03-03161-4  
42 74.5 7.3 490 6 US-10-350-385-4  
43 74.5 7.3 492 1 PCT-US03-03161-7  
44 74.5 7.3 492 6 US-10-350-385-7  
45 74.5 7.3 538 5 US-09-724-676-74895

#### ALIGNMENTS

#### RESULT 1

US-10-282-122A-56549  
; Sequence 56549, Application US/10282122A

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282.122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 56549

; LENGTH: 931

; TYPE: PRT

; ORGANISM: Escherichia coli

; US-10-282-122A-56549

```
Query Match          9.5% Score 97; DB 6; Length 931;
Best Local Similarity 27.3%; Pred.No.0.84;
Matches 38; Conservative 22; Mismatches 55; Indels 24; Gaps 5;

Qy 1 MEITSSKFLILLTATSSLTSNIFCA-----DELVMNLHSENDKYSEPRGYPKGER 55
      |||| | ||| : || | : ::| : | | : | | : | | : 
Db 1 MEIMRNLCFLTLLVATLPLGRILAAALPDKELTGQLDNGRLMYI--PHAHPKDQV 58
      : | : | : | : | : | : | : | : | : | : | : | : 

Qy 56 SLNF-----EELKWGPKNVIK-----MSTPAVNKMPSFANPLRFRGNVOEERS 101
      : | : | : | : | : | : | : | : | : | : | : | : 
Db 59 NLWLQIHGTSLQEDNELGVAHFVEHHMFNGTKTWPGNKVIETTESMGRLRGFDNVAYTS 118
      : | : | : | : | : | : | : | : | : | : | : | : 

Qy 102 AGAT---ANPLRSRNMWE 117
       : | : | : | : | : | : | : | : | : | : | : | : 
Db 119 YDETVYQVSLEPTTKOKNIQ 137
        : | : | : | : | : | : | : | : | : | : | : | : 

RESULT 2
US-10-369-493-2002
: Sequence 202, Application US/10369493
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkie, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: PRIOR FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 2002
: LENGTH: 629
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-369-493-2002

Query Match          8.3% Score 84; DB 6; Length 629;
Best Local Similarity 21.7%; Pred.No.9.5;
Matches 34; Conservative 28; Mismatches 51; Indels 44; Gaps 6;

Qy 52 KGPSLNFEELKGWPKNTMKMSPAVNKMPSFANPLRFRGNVOEERSACATANPLR 111
      : | : | : | | : | : | : | : | : | : | : | : | : 
Db 342 EADRLLN-QSFGWC PKMLSHLKTDLDLTLP-----GNVKMIFSATLTTNTEKL 390
      : | : | : | : | : | : | : | : | : | : | : | : | : 

Qy 112 SGRNM-----EVSLVRVPNLPORFG-RITTAKSVCR----MLS DLCQGSMHSPCAND 159
      : | : | : | : | : | : | : | : | : | : | : | : | : 
Db 391 NGUNLYKPRLF LKQT DKLYOLPNKLFNEFNINPTAKSVYKPLILLYSIQCFAHSPAIA-- 448
      : | : | : | : | : | : | : | : | : | : | : | : | : 

Qy 160 LFYSMTCHQEIQNPQDGSRLLFKKIDDAELKOEK 196
      : | : | : | : | : | : | : | : | : | : | : | : | : 
Db 449 -----AKLI IFVKSNESSIRLSK 466
        : | : | : | : | : | : | : | : | : | : | : | : | : 

RESULT 3
US-10-218-140-3988
: Sequence 3988, Application US/10218140
: GENERAL INFORMATION:
: APPLICANT: Leach, Martin D.
: APPLICANT: Shinkets, Richard A.
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES
: FILE REFERENCE: ENCODED THEREBY
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES
: FILE REFERENCE: 15966-543 CON
: CURRENT APPLICATION NUMBER: US/10/218,140
: CURRENT FILING DATE: 2002-08-12
: PRIOR APPLICATION NUMBER: 09/540,763
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: 60/127,728
: PRIOR FILING DATE: 1999-04-05
: PRIOR APPLICATION NUMBER: 60/127,636
```



Matches 50; Conservative 44; Mismatches 79; Indels 64; Gaps 13;

QY 10 ILTLATSSLLTSNIP--CADELVMSN--LHSKE-----NYDK-----YSEPRGYPKGER 55  
Db 13 LTRMAMDSSLOARLFPGLAIKIQRSNGLIHSANVRTVNLEKSCSVSEWAE--GGATKG-K 70

QY 56 SLNFEELKDWGPKNVKIMSTPAVNKMP-HSFANLPLRFGNRVVOERSAGATANLPL---- 110  
Db 71 EIDFDDVAAINPE-----LLQLLPLHPKTNLPLQENVTIQKRRSVNSKIPAPKES 122

QY 111 ---RSGRNMESVLR-----RVNPLPQRFGRTRTTAKSVCRMLS 145  
Db 123 LRSRSTRMSTVSELRTITAQENDMEVELPAAANSRKQFSVPPATRPSPCAVAEIPLRMVS 182

QY 146 DLCOGSMH-----SPCANDL--FYSMTCOHQEIQNPQOKSRRLLFKKIDDAELKQEK 196  
Db 183 EEMEEQVHSIRGSSSANPVNSVRRKSCLVKEVEKMKNKREE----KKAQNSEMRMKR 235

RESULT 9  
US-10-017-161-2042  
; Sequence 2042, Application US/10017161  
; GENERAL INFORMATION:  
; APPLICANT: SUWA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABURATANI, HIROYUKI  
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 084335/0152  
; CURRENT APPLICATION NUMBER: US/10/017,161  
; PRIOR FILING DATE: 2002-12-18  
; PRIOR APPLICATION NUMBER: JP 2001/246789  
; FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2430  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 2042  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-017-161-2042

Query Match 8.0%; Score 81; DB 6; Length 502;  
Best Local Similarity 31.1%; Pred. No. 14;  
Matches 36; Conservative 15; Mismatches 49; Indels 20; Gaps 7;

QY 6 SKLFILLATSSLLTSNIPFCADLVMSNL-----HSKENYDKYSEPRGYPKGER 55  
Db 358 SGLF-LTVG--SSILTA--FLLSQLVORKLDQKTEICRTHSRDRPKRLVGRGTAKGAR 413

QY 56 SLNFEELKDWGPKNVKIMSTPAVNKMP-HSFANLPLRFGNRVVOERSAGATANLPLRSGRN 115  
Db 414 AL---PGRGWPRGRSRSPSVG--ARSERRP-RFWRSLRRREPAPASCRLLLSGHE 467

QY 116 ME 117  
Db 468 AD 469

RESULT 10  
US-09-724-676-74897  
; Sequence 74897, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Comugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; NUMBER OF SEQ ID NOS: 2000-11-28  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 74897  
; LENGTH: 537  
; TYPE: PRT  
; ORGANISM: Homo sapiens

Matches 50; Conservative 44; Mismatches 79; Indels 64; Gaps 13;

QY 10 ILTLATSSLLTSNIP--CADELVMSN--LHSKE-----NYDK-----YSEPRGYPKGER 55  
Db 13 LTRMAMDSSLOARLFPGLAIKIQRSNGLIHSANVRTVNLEKSCSVSEWAE--GGATKG-K 70

QY 56 SLNFEELKDWGPKNVKIMSTPAVNKMP-HSFANLPLRFGNRVVOERSAGATANLPL---- 110  
Db 71 EIDFDDVAAINPE-----LLQLLPLHPKTNLPLQENVTIQKRRSVNSKIPAPKES 122

QY 111 ---RSGRNMESVLR-----RVNPLPQRFGRTRTTAKSVCRMLS 145  
Db 123 LRSRSTRMSTVSELRTITAQENDMEVELPAAANSRKQFSVPPATRPSPCAVAEIPLRMVS 182

QY 146 DLCOGSMH-----SPCANDL--FYSMTCOHQEIQNPQOKSRRLLFKKIDDAELKQEK 196  
Db 183 EEMEEQVHSIRGSSSANPVNSVRRKSCLVKEVEKMKNKREE----KKAQNSEMRMKR 235

RESULT 11  
US-09-724-676A-74897  
; Sequence 74897, Application US/09724676A  
; GENERAL INFORMATION:  
; APPLICANT: Comugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 74897  
; LENGTH: 537  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676A-74897

Query Match 8.0%; Score 81; DB 5; Length 537;  
Best Local Similarity 21.1%; Pred. No. 16;  
Matches 50; Conservative 44; Mismatches 79; Indels 64; Gaps 13;

QY 10 ILTLATSSLLTSNIP--CADELVMSN--LHSKE-----NYDK-----YSEPRGYPKGER 55  
Db 13 LTRMAMDSSLOARLFPGLAIKIQRSNGLIHSANVRTVNLEKSCSVSEWAE--GGATKG-K 70

QY 56 SLNFEELKDWGPKNVKIMSTPAVNKMP-HSFANLPLRFGNRVVOERSAGATANLPL---- 110  
Db 71 EIDFDDVAAINPE-----LLQLLPLHPKTNLPLQENVTIQKRRSVNSKIPAPKES 122

QY 111 ---RSGRNMESVLR-----RVNPLPQRFGRTRTTAKSVCRMLS 145  
Db 123 LRSRSTRMSTVSELRTITAQENDMEVELPAAANSRKQFSVPPATRPSPCAVAEIPLRMVS 182

QY 146 DLCOGSMH-----SPCANDL--FYSMTCOHQEIQNPQOKSRRLLFKKIDDAELKQEK 196  
Db 183 EEMEEQVHSIRGSSSANPVNSVRRKSCLVKEVEKMKNKREE----KKAQNSEMRMKR 235

RESULT 12  
US-09-724-676-74919  
; Sequence 74919, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Comugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 74919

```
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 74922
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-74922

Query Match      8.0%; Score 81; DB 5; Length 713;
Best Local Similarity 21.1%; Pred. No. 21;
Matches 50; Conservative 44; Mismatches 79; Indels 64; Gaps 13;

QY   10 ILLTATSSLLTSNIF--CADELVMSN---LHSKE-----NYDK-----YSEPRGYPKGER 55
    :|:|:|:|:|:~::~||~::||~::||~::||~::||~::||~::||~::||~::||~::||~::
Db    13 LTLRMADSSLSQARLFPGGLAIKQRSNGLIHSANVRTVNLEKSCVSVEWAE--GGATKG-K 70
    :::::::|:~::||~::||~::||~::||~::||~::||~::||~::||~::||~::

QY   56 SLNFPEELKDGPKNVIKMSTPAVNKP-HSFANLPRLFRGNRNVQEERSAGATANLPL---- 110
    :::::::|:~::||~::||~::||~::||~::||~::||~::||~::||~::||~::
Db    71 EIDFDVAAINPE-----LLQLPLHPKTNLPLEQENVTIQOKRRSVNSKIAPAKES 122
    :::::::|:~::||~::||~::||~::||~::||~::||~::||~::||~::||~::

QY   111 ---RSGRNMVESLVLR-----RVPNLPORFGRTTTAAKSVCRLMS 145
     || ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~||
Db    123 LRSRSTRMSTVSELRIITAQENDMEVELPAAANSRKQFSVPPTPPSPCPAVAEIPLRWVS 182
          |~::~||~::||~::||~::||~::||~::||~::||~::||~::||~::||~::

QY   146 DLCOGSMH-----SPCANDL--FYSNTCQHQTIONDPOKRSLFLFKIDDIAELKOEK 196
    :|:|:|:|:|:~::~||~::||~::||~::||~::||~::||~::||~::||~::||~::
Db    183 EMEEQVHSIRGSSSANNPVNSVRKSLCKVEKMKNKREE---KKAQNSEMNRKR 235
           :|:|:|:~::~||~::||~::||~::||~::||~::||~::||~::||~::||~::

RESULT 15
US-09-724-676A-74922
; Sequence 74922, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 74922
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-74922

Query Match      8.0%; Score 81; DB 5; Length 713;
Best Local Similarity 21.1%; Pred. No. 21;
Matches 50; Conservative 44; Mismatches 79; Indels 64; Gaps 13;

QY   10 ILLTATSSLLTSNIF--CADELVMSN---LHSKE-----NYDK-----YSEPRGYPKGER 55
    :|:|:|:|:|:~::~||~::||~::||~::||~::||~::||~::||~::||~::||~::
Db    13 LTLRMADSSLSQARLFPGGLAIKQRSNGLIHSANVRTVNLEKSCVSVEWAE--GGATKG-K 70
    :::::::|:~::||~::||~::||~::||~::||~::||~::||~::||~::||~::

QY   56 SLNFPEELKDGPKNVIKMSTPAVNKP-HSFANLPRLFRGNRNVQEERSAGATANLPL---- 110
    :::::::|:~::||~::||~::||~::||~::||~::||~::||~::||~::||~::
Db    71 EIDFDVAAINPE-----LLQLPLHPKTNLPLEQENVTIQOKRRSVNSKIAPAKES 122
    :::::::|:~::||~::||~::||~::||~::||~::||~::||~::||~::||~::

QY   111 ---RSGRNMVESLVLR-----RVPNLPORFGRTTTAAKSVCRLMS 145
     || ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~|| ~||
Db    123 LRSRSTRMSTVSELRIITAQENDMEVELPAAANSRKQFSVPPTPPSPCPAVAEIPLRWVS 182
          |~::~||~::||~::||~::||~::||~::||~::||~::||~::||~::||~::

QY   146 DLCOGSMH-----SPCANDL--FYSNTCQHQTIONDPOKRSLFLFKIDDIAELKOEK 196
    :|:|:|:~::~||~::||~::||~::||~::||~::||~::||~::||~::||~::
Db    183 EMEEQVHSIRGSSSANNPVNSVRKSLCKVEKMKNKREE---KKAQNSEMNRKR 235
           :|:|:|:~::~||~::||~::||~::||~::||~::||~::||~::||~::||~::
```

```

; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-74919

Query Match      8.0%; Score 81; DB 5; Length 709;
Best Local Similarity 21.1%; Pred. No. 21;
Matches 50; Conservative 44; Mismatches 79; Indels 64; Gaps 13;

QY 10 ILLTATSSLLTSNIP--CADELVMNSN--LHSKE-----NYDK-----YSEPRGYPKGER 55
      : | | | | | : | : | : | | | | | : | | | : | | | :
DB 13 LTLRMAMDSSLLQARLFPGLAIKQRNGLIHSANVRTVNLEKSCVSVEWAE--GGATKG-K 70
      : | | | | | : | : | : | | | | | : | | | : | | | :
QY 56 SLNFEELKDWGPKNVIKMSTPAVNKMP--HSFANLPLRFGRNVQEERSAGATANLPL---- 110
      : | | | | | : | : | : | | | | | : | | | : | | | :
DB 71 EIDFDVAAINPE-----LLQLLPHLPTKNLPLQENVTIQKKRRSVNSKIPAPKES 122
      : | | | | | : | : | : | | | | | : | | | : | | | :
QY 111 ---RSGRNMEVSLVR-----RVPNLPRQFRGRTTTAKSVCRMLS 145
      : | | | | | : | : | : | | | | | : | | | : | | | :
DB 123 LRSRSTRMSTVSELRTAQENDMEVELPAAANSRRKFQSVPPAPTRPSCPVAVEIPLRMVS 182
      : | | | | | : | : | : | | | | | : | | | : | | | :
QY 146 DLCOGSMH-----SPCANDL--FYSMTCHOQEIQNDQKSRLLFKKIDDAELKQEK 196
      : | | | | | : | : | : | | | | | : | | | : | | | :
DB 183 EEMEEQVHSIRGSSSANDPVNSVRRKSCLVKEVEKMKNKREE---KKAQNSEMMRKR 235
      : | | | | | : | : | : | | | | | : | | | : | | | :

RESULT 13
US-09-724-676A-74919
; Sequence 74919, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 74919
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-74919

Query Match      8.0%; Score 81; DB 5; Length 709;
Best Local Similarity 21.1%; Pred. No. 21;
Matches 50; Conservative 44; Mismatches 79; Indels 64; Gaps 13;

QY 10 ILLTATSSLLTSNIP--CADELVMNSN--LHSKE-----NYDK-----YSEPRGYPKGER 55
      : | | | | | : | : | : | | | | | : | | | : | | | :
DB 13 LTLRMAMDSSLLQARLFPGLAIKQRNGLIHSANVRTVNLEKSCVSVEWAE--GGATKG-K 70
      : | | | | | : | : | : | | | | | : | | | : | | | :
QY 56 SLNFEELKDWGPKNVIKMSTPAVNKMP--HSFANLPLRFGRNVQEERSAGATANLPL---- 110
      : | | | | | : | : | : | | | | | : | | | : | | | :
DB 71 EIDFDVAAINPE-----LLQLLPHLPTKNLPLQENVTIQKKRRSVNSKIPAPKES 122
      : | | | | | : | : | : | | | | | : | | | : | | | :
QY 111 ---RSGRNMEVSLVR-----RVPNLPRQFRGRTTTAKSVCRMLS 145
      : | | | | | : | : | : | | | | | : | | | : | | | :
DB 123 LRSRSTRMSTVSELRTAQENDMEVELPAAANSRRKFQSVPPAPTRPSCPVAVEIPLRMVS 182
      : | | | | | : | : | : | | | | | : | | | : | | | :
QY 146 DLCOGSMH-----SPCANDL--FYSMTCHOQEIQNDQKSRLLFKKIDDAELKQEK 196
      : | | | | | : | : | : | | | | | : | | | : | | | :
DB 183 EEMEEQVHSIRGSSSANDPVNSVRRKSCLVKEVEKMKNKREE---KKAQNSEMMRKR 235
      : | | | | | : | : | : | | | | | : | | | : | | | :

RESULT 14
US-09-724-676-74922
; Sequence 74922, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28

```

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: March 26, 2003, 10:12:39 ; Search time 43 Seconds  
(without alignments)  
438.194 Million cell updates/sec

Title: US-09-831-758A-8  
Perfect score: 1018  
Sequence: 1 MEIISKLFIILLTATSSLL.....KQSRLLFKKIDDAELKQEK 196

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*

- 1: pir1.\*
- 2: pir2.\*
- 3: pir3.\*
- 4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	98	9.6	426	C96804	hypothetical prote
2	97	9.5	931	A64903	probable zinc prot
3	96	9.4	931	C90891	probable peptidase
4	96	9.4	931	F85726	probable peptidase
5	86.5	8.5	1211	D64116	exodeoxyribonuclea
6	85.5	8.4	1015	T32984	hypothetical prote
7	84	8.3	629	S63369	hypothetical prote
8	83.5	8.2	1163	J50366	tight junction pro
9	82	8.1	261	T25206	hypothetical prote
10	81.5	8.0	1358	A29360	SIR4 protein - yea
11	81	8.0	1343	AF0611	cell division prot
12	80	7.9	1084	S19661	DNA-directed DNA p
13	80	7.9	1086	T43266	DNA-directed DNA p
14	80	7.9	1086	T43266	DNA polymerase del
15	80	7.9	1305	A80168	probable cell divi
16	79.5	7.8	144	G90573	50S ribosomal prot
17	79.5	7.8	327	D95408	probable oxidoredu
18	79.5	7.8	449	QADA5	maturation (p1Va2)
19	79.5	7.8	449	QADA2	maturation (p1Va2)
20	79.5	7.8	1275	T41523	hypothetical rhoI
21	79.5	7.8	1957	A45627	myosin heavy chain
22	79.5	7.8	609	S55957	hypothetical prote
23	78.5	7.7	688	T13248	NADH2 dehydrogenas
24	78	7.7	1204	T19918	hypothetical prote
25	77	7.6	448	QADA7	maturation (p1Va2)
26	77	7.6	579	H87451	oxidoreductase, GM
27	77	7.6	1051	E83330	RND divalent metal
28	76.5	7.5	420	A37343	keratin 13, type I
29	76.5	7.5	458	KRHU3	keratin 13, type I

30	76.5	7.5	473	2	H84550	probable obtusifol
31	76.5	7.5	1357	2	T16860	hypothetical prote
32	76	7.5	1436	2	B81704	conserved hypothet
33	76	7.5	1893	2	T22661	hypothetical prote
34	75.5	7.4	361	2	T40857	hypothetical prote
35	75.5	7.4	371	2	T13023	drought-inducible
36	75.5	7.4	1642	2	T19130	hypothetical prote
37	75	7.4	517	2	D86637	ABC transporter AT
38	75	7.4	551	2	T03793	calmodulin-binding
39	75	7.4	582	2	H95950	probable ABC trans
40	74.5	7.3	492	2	T14820	obtusifolios 14-al
41	74.5	7.3	943	2	S28400	gag-like protein -
42	74.5	7.3	1073	2	T33764	hypothetical prote
43	74.5	7.3	1683	2	T30885	complement compone
44	74.5	7.3	1963	1	MWKW	myosin heavy chain
45	74	7.3	688	2	T13253	NADH2 dehydrogenas

## ALIGNMENTS

### RESULT 1

C96804  
hypothetical protein T5M16.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 24-Aug-2001  
C:Accession: C96804  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Jansen, N.F.; Hughes, B.; Conn, L.; Conway, A.B.; Huizler, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Ma  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: C96804  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-426 <STO>  
A:Cross-references: GB:AE005173; NID:G6382497; PIDN:AAF07783.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T5M16.8  
A:Map position: 1  
C:Superfamily: cytochrome-c peroxidase

Query Match 9.6%; Score 98; DB 2; Length 426;  
Best Local Similarity 29.3%; Pred. No. 0.16;  
Matches 53; Conservative 20; Mismatches 66; Indels 42; Gaps 11;

Qy	2	ETISKRLILLTATSSLLTSNIFCADELVNSLNHSEKNDKYSEP---RGYPKERSLN 58
Db	85	QLISAKEDIKVLRTK-----FCHPLVLRLGMDAGTYNKNIIEWPUGGANG--SUR 135
Qy	59	RE-ELK---DWGPKNVKIMSTPAVNMPIH-SPANPLRFRNVQBERSAGATANPLRSG 113
Db	136	FBAELKHANAGLLNALKLIQLKDKYPMISYADL-----FQLASATA-IEEAGG 184
Qy	114	RNMEYSLVRRVNPPLRFQRT--TTAKSVCRMLSDLCQSMHSPC--ANDLFVMTCHQHE 170
Db	185	-----PDIPMKYGRVDVWAPEQCPEGRPLDAGPPSPADHLRDVFRMGLDDKE 233
Qy	171	I 171
Db	234	I 234

### RESULT 2

A64903  
probable zinc proteinase yddC (EC 3.4.99.-) - Escherichia coli (strain K-12)  
C:Species: Escherichia coli

probable peptidase ppgL [imported] - Escherichia coli (strain O157:H7, substrain EDL\_933)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: F85726  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.;  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.;  
Rabinovitch, A.; Brinkley, S.; Fisk, S.; Moreno, H.; Weng, S.; White, O.; Peterson, J.;  
NATURE 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: F85726  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-931 <STO>  
A:Cross-references: GB:AE005174; NTD:gl2515180; PIDN:AAG56274.1; GSPDB:GN00145; UWGCE:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ppgL

[illegible]

RESULT 5

D64116  
exodeoxyribonuclease V 135K chain homolog - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: D64116  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.H.;  
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidm,  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Ven  
A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: D64116  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1211 <TIGR>  
A:Cross-references: GB:U32811; GB:L42023; NID:g1574774; PID:g1574774  
C:Superfamily: exodeoxyribonuclease V 135K chain

Query Match	8.5%; Score 86.5; DB 2; Length 1211;
Best Local Similarity	20.9%; Pred. No. 7.4;
Matches	39; Conservative 33; Mismatches 80; Indels 35; Gaps 7;
QY	11 LLTLATSSLLTGN-IFCADELV-----MSNLHSENYDKYSEPRGYPKGER 55
DB	900 LEALPTLSINTKDDPKASEFTGNIQDWRTSFTSTIEQAHRQNY--FTE-----SAGKK 953
QY	56 SLNFEEKDWGPKNVLKMTTPAVNKKPHGFANPL-----RFGRNVOERSAGATA 106
DB	954 HAVFDDAKDYDSQNALEISTALNENSNILDLPRGQVGQTALRRHFENCYFSDLANTEE 1013
QY	107 NLPLSRGRNVEVLRRVNLPRFGRTTTAKSVCRMLSLDCOGSMHSPCANDL-FYSMT 165
DB	1014 IDKLRSLOLDEFTFTESLQNWLOOIHSHTPLSNIGTALADLA-----NRKCIKEMPFYLAII 1069
QY	166 CQHQEIQ 172

QY 166 CQHQEIQ 172  
:1 ::

Db 1070 REHFDVE 1076

RESULT 6

T32984  
hypothetical protein K02D7.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32984  
R:Du, Z.; Magg1, L.  
submitted to the EMBL Data Library, February 1998  
A:Description: The sequence of C. elegans cosmid K02D7.  
A:Reference number: Z21259  
A:Accession: T32984  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1015 <DU2>  
A:Cross-references: EMBL:AF045645; PIDN:AAC02607.1; GSPDB:GN00022; CESP:K02D7.4  
A:Experimental source: strain Bristol N2; clone K02D7  
C:Genetics:  
A:Gene: CESP:K02D7.4  
A:Map position: 4  
A:Introns: 51/1; 103/1; 188/3; 304/2; 339/2; 392/3; 448/2; 545/3; 646/3; 794/3; 910/2

Query Match 8.4%; Score 85.5; DB 2; Length 1015;  
Best Local Similarity 19.8%; Pred. No. 7.4; Indels 81; Gaps 8;  
Matches 52; Conservative 35; Mismatches 94; Indels 81; Gaps 8;  
QY 12 LTATSSLLTSNIFCADELVMSNLHSENYDKYSEPRGYPKGRSLNFEELKDWGPKNVI 71  
DB 464 ITTAREVLLTASPDYLDLLFGISQSSSNNEKWHKQLMYGLSLDKSEY--WKVANTI 521  
QY 72 -----KMSTPAVN-----KMFHFAANLPL-----R 91  
DB 522 ATVLNKRCEASTSSLNKCKRETIVNKFITDLTAGGVEVRVLEVINIPIFGSTYFAKK 581  
QY 92 FGRRVQERSAGATANLPLRSGRNNEVSLVRVNPQLPQRFGRRTTAKSVCRMLSLDLCQ-- 149  
DB 582 FICETESEDVQKALNVILAAKSNLYETQVSGLYEQEEMRGYSAAHAEFFRHIS-ICQPK 640  
QY 150 -----GSMHSPCANDLFYSWTCQHQ-----EIQNPDQK 177  
DB 641 KKIGAPLTHKLKLFRTNCSQETPTSHSQLAIDLLKCPVDHQNVATLILRTLTLPDQO 700  
QY 178 QSRRLFKKIDDA----ELKQE 195  
DB 701 EKWHYLYKAIEASGNKDELKAE 722

RESULT 7

S63369  
hypothetical protein YNR038w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein N3302  
C:Species: Saccharomyces cerevisiae  
C:Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 02-Feb-2001  
C:Accession: S63369  
R:Pohl, T.M.  
submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63346  
A:Accession: S63369  
A:Molecule type: DNA  
A:Residues: 1-629 <POH>  
A:Cross-references: EMBL:Z71653; NID:g1302540; PIDN:CAA96318.1; PID:e239589; PID:g130254  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:DBP6  
A:Cross-references: SGD:S0005321; MIPS:YNR038w  
A:Map position: 14R  
C:Superfamily: fruit fly gene Dbp73D protein  
C:Keywords: ATP; nucleotide binding; P-loop  
F;234-241/Region: nucleotide-binding motif A (P-loop)  
F;337-342/Region: nucleotide-binding motif B  
F;341-344/Region: DEAD motif

Query Match 8.3%; Score 84; DB 2; Length 629;  
Best Local Similarity 21.7%; Pred. No. 5.7; Indels 44; Gaps 6;  
Matches 34; Conservative 28; Mismatches 51; Indels 44; Gaps 6;

QY 52 KGRSLNFEELKDWGPKNVIKMTSPAVNKPMPHSFANLPLRFRGNVQERSAGATANLPLR 111  
DB 342 EADRLN-QSFGQWCPKMLSHLTKDLDTP-----GNVKMFISATLTTNTEKL 390  
QY 112 SGRNM-----EVSILVRVNPQLPQRFGR--RTTAKSVCR---MLSDLCGSMHSPCAND 159  
DB 391 NGLNLYRKLFLKQTDKLYQLPNKLNFEFNINPTAKSVYKPLLLYLLYSICOFMAHSPIA-- 448  
QY 160 LFYSMTCHQOEIQNDOKQSRRLIFKKIDDAELKQEK 196  
DB 449 -----AKILIFVKSNESSIRLSK 466

RESULT 8

JE0366  
tight junction protein, ZO-2 - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: JE0366  
R:Collins, J.R.; Rizzolo, L.J.  
Biochem. Biophys. Res. Commun. 252, 617-622, 1998  
A:Title: Protein-binding domains of the tight junction protein, ZO-2, are highly  
A:Reference number: JE0366; MUID:99057550; PMID:9837755  
A:Accession: JE0366  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1163 <COL>  
A:Cross-references: GB:AF085184; NID:g3820579; PIDN:AAC95469.1; PID:g3820580

Query Match 8.2%; Score 83.5; DB 2; Length 1163;  
Best Local Similarity 27.6%; Pred. No. 14; Indels 19; Gaps 3;  
Matches 24; Conservative 17; Mismatches 27; Indels 19; Gaps 3;

QY 50 YPKGRSLNFEELKDWGPKN-----VTKMSTPAVNKPMPHSFANLPLRFRGNVQER 100  
DB 34 FENGETSIVISDLVPGGPADGLLQENDRVIVNGTPEMEN-VPHSFA-----VQQLR 83  
QY 101 SAGATANLPLRSGRNNEVSLVRVNPQL 127  
DB 84 KSGKVATIVVKKRPKVQAAALRKNPSL 110

RESULT 9

T23206  
hypothetical protein T23G7.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T23206  
R:Barlow, K.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z19995  
A:Accession: T23206  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-261 <WIL>  
A:Cross-references: EMBL:Z68319; PIDN:CAA9703.1; GSPDB:GN00020; CESP:T23G7.5  
A:Experimental source: clone T23G7  
C:Genetics:  
A:Gene: CESP:T23G7.5  
A:Map position: 2  
A:Introns: 56/2; 137/2; 170/1; 207/1

Query Match 8.1%; Score 82; DB 2; Length 261;  
Best Local Similarity 26.0%; Pred. No. 3; Indels 22; Gaps 7;  
Matches 38; Conservative 23; Mismatches 63; Indels 22; Gaps 7;

QY 12 LTATSSLLTSNIFCADELVMSNLHSENYDKYSEPRGYPK--GERSLNFEELKDWGPKN 69  
DB 341-344/Region: nucleotide-binding motif B

Db 13 LNLSHSLI--SIFSVS--VMSYHHNNHQ--HRPRGYERLPGLKRLPDRWNIYDNGRD 66  
QY 70 V-----IKMSTPAVNKMPHSFANLPLFRGNVOERSAGATANPLRSGRNMEVSLVRV 124  
Db 67 IDGTRFVPPKTP-----LDSSFFD-----GKNMPELVQFGVKTLISLAQQAQKQIGLVIDL 117  
QY 125 PNLPPQFGRTRTTAKSVCRMLSDLCQG 150  
Db 118 TNTDRIYKKTENADHGKVKLKLNCPG 143  
RESULT 10  
A29360  
SIR4 protein - yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein YD934.12; protein YDR227w; STE9 protein  
C;Species: Saccharomyces cerevisiae  
C;Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 21-Jul-2000  
C;Accession: A29360; S47935; S59434; S53988; S47476; S47950  
R;Marshall, M.; Mahoney, D.; Rose, A.; Hicks, J.B.; Broach, J.R.  
Mol. Cell. Biol. 7, 4441-4452, 1987  
A;Title: Functional domains of SIR4, a gene required for position effect regulation in S  
A;Reference number: A29360; MUID:88142836; PMID:3325825  
A;Accession: A29360  
A;Molecule type: DNA  
A;Residues: 1-1358 <MAR>  
A;Cross-references: GB:M37249; NID:g531115; PIDN:AAA20881.1; PID:g531116  
R;Davies, C.J.; Hutchison III, C.A.  
submitted to the EMBL Data Library, September 1994  
A;Description: Tn3 transposon/deletion sequencing of a 9.4kb DNA fragment: Characterisat  
A;Reference number: S47932  
A;Accession: S47935  
A;Molecule type: DNA  
A;Residues: 1-993, 'L', 995-1358 <DAV>  
A;Cross-references: EMBL:L35344; EMBL:U13239; EMBL:Z36548  
R;Murphy, L.; Harris, D.  
submitted to the EMBL Data Library, March 1995  
A;Reference number: S59423  
A;Accession: S59434  
A;Molecule type: DNA  
A;Residues: 1-1358 <MUR>  
A;Cross-references: EMBL:Z48612; NID:g728671; PID:g728683; MIPS:YDR227w  
A;Experimental source: strain AB972  
R;Davies, C.J.; Hutchison III, C.A.  
Nucleic Acids Res. 23, 507-514, 1995  
A;Title: Insertion site specificity of the transposon Tn3.  
A;Reference number: S53985; MUID:95192063; PMID:7885847  
A;Accession: S53988  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-993, 'L', 995-1358 <DAW>  
A;Cross-references: EMBL:U13239; NID:g532747; PIDN:ANC33144.1; PID:g532751  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
C;Genetics:  
A;Gene: SGD: SIR4; STE9  
A;Cross-references: SGD:S0002635; MIPS:YDR227w  
A;Map position: 4R  
C;Keywords: DNA binding; transcription regulation  
Query Match 8.0%; Score 81.5; DB 2; Length 1358;  
Best Local Similarity 23.2%; Pred. No. 26;  
Matches 48; Conservative 34; Mismatches 84; Indels 41; Gaps 9;  
QY 7 KLFILLIATSSLTNIFCADELVMNL-----HSKENYDKYSEPRGYPKGRSLNFEEL 62  
Db 880 RLOYLPLAVSTFNTNTEKNDVTNKNINIGHSQBNSSAKPSQIPTVSPGLGEET 939  
QY 63 KDWGPKNVKMTSPAVNKMPSFANLPLFRGNVOERSAGA-----TANLPLRSGRNME 117  
Db 940 K-----LSTPTKSNRVSHSDTNSK--PKNTKENLSKSSRWQEWLANLKL-----IS 986  
QY 118 VSLVRVNPPLPQFGRTRTTAKSVCRMLSDLCQGSMSHSPCANDL-----FYSM 164  
Db 987 VSLVDEFPSELSDSRQIINEKM-QLLKDI FANLKLKSAISNNFRESDIILKGEIEDYPM 1045

QY 165 TCQ-----HOEIQN-PDKOSRRLLFKK 186  
Db 1046 SSEIKIYYNELQKPKDAKAREWSEMK 1072

## RESULT 11

AF0611  
cell division protein FtsK [imported] - Salmonella enterica subsp. enterica serovar  
C;Species: Salmonella enterica subsp. enterica serovar typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C;Accession: AF0611  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Chur  
th, T.; Connerthon, P.; Cronin, A.; Davis, R.M.; Dowd, L.; White, N.; Fal  
, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Ouail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ;  
A;Reference number: AB0502; PMID:11677608  
A;Accession: AF0611  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1343 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD05360.1; PID:g16502124; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY0958

Query Match 8.0%; Score 81; DB 2; Length 1343;  
Best Local Similarity 25.2%; Pred. No. 29;  
Matches 29; Conservative 24; Mismatches 42; Indels 20; Gaps 6;

QY 65 WGPKNVVKMTSPAVNKMPSFANLPLFRGNVOE-----SPANLPLFRGNVOE-----ERSAGATANPLRS 112

Db 1111 WKPQDSMDVQHPVLEKLPYIVLVLDEPADLMVTGKKEELIARLAKAARAGTHLVAT 1170

QY 113 GR---NNEVSLVRVNPPLPQFGRTRTTAKSVCRMLSDLCQGSMSHSPCA-NDLFYS 163

Db 1171 QRPSSVDVITGLIK--ANIPTRIAFTVSSKIDSRITLD--OGGAESLLGMGDMLYS 1221

## RESULT 12

S19661  
DNA-directed DNA polymerase (EC 2.7.7.7) III large chain - fission yeast (Schizosac  
C;Species: Schizosaccharomyces pombe  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: S19661  
R;Pignade, G.; Bouvier, D.; de Recondo, A.M.; Baldacchi, G.  
J. Mol. Biol. 222, 209-218, 1991  
A;Title: Characterization of the POL3 gene product from Schizosaccharomyces pombe 1r  
A;Reference number: S19661; MUID:92071954; PMID:1960723  
A;Accession: S19661  
A;Molecule type: DNA  
A;Residues: 1-1084 <PIG>  
A;Cross-references: EMBL:X59278; NID:g5010; PIDN:CAA41968.1; PID:g5011  
C;Genetics:  
A;Introns: 77/1  
C;Superfamily: herpesvirus DNA-directed DNA polymerase  
C;Keywords: DNA binding; nucleotidyltransferase; nucleus

Query Match 7.9%; Score 80; DB 1; Length 1084;

Best Local Similarity 27.9%; Pred. No. 27;

Matches 38; Conservative 20; Mismatches 36; Indels 42; Gaps 11;

QY 71 IKMSTPAVNKMPHSFANLPLFRGNVOERSAGATANPLRSGR-----NNEVSLVR 122

Db 971 ISMAAPSVGGI-----MKFA--VRVETCLGCKA--PIKKGKLTALCENCLNSAEIYQ 1018

QY 123 R----VNPPLPQFGRTRTTAKSVCRMLSDLCQGSMSHSP--CAN---DLFYSMTCCOHOEIQN 173

Db 1019 RQVAQVNDLLEVRFARLWTQ---COR----CGSMHQMDVICTSRDCCPIFYMRIAHEKKLQ- 1070

QY 174 PDKOSRRLLFKKITDD 189

Db 1071 ----QSVDLL-KRFE 1081

RESULT 13

T43266  
DNA-directed DNA polymerase (EC 2.7.7.7) delta chain - fission yeast (Schizosaccharomyces pombe)  
N:Alternate names: DNA polymerase delta  
C:Species: Schizosaccharomyces pombe  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T43266  
R:Pignede, G.; Bouvier, D.; de Recondo, A.M.; Baldacci, G.  
J. Mol. Biol. 222, 209-218, 1991  
A:Title: Characterization of the POL3 gene product from Schizosaccharomyces pombe indica  
A:Reference number: S19661; MUID:92071954; PMID:1960723  
A:Accession: T43266  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1086 <PIG>  
A:Cross-references: EMBL:L07734; MID:q173383; PIDN:AAA35303.1; PID:q173384  
C:Genetics:  
A:Gene: pold  
A:Introns: 77/1  
C:Superfamily: herpesvirus DNA-directed DNA polymerase  
C:Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 7.9%; Score 80; DB 2; Length 1086;  
Best Local Similarity 27.9%; Pred. No. 27;  
Matches 38; Conservative 20; Mismatches 36; Indels 42; Gaps 11;

QY 71 IKMSTPAVKNMPSFANLPLRFRGNVQERSAGATANLPLRSGR-----NMEVSLVR 122  
Db 973 ISMAAPSVGGI-----MKFA--VKVETCLGCKA--PIKKGKTALENCNLSAEYQ 1020  
QY 123 R----VPNLPQRFGRTRTTAKSVCRMLSDLCQGSMHSP--CAN---DLFYSMTQHOBIQN 173  
Db 1021 ROVAQVNDLEVRFARLWTQ---CQR-----COGSMHQDVICTSRDCPIFYMRIAEHKKLQ- 1072  
QY 174 PDQKQSRRLFLFKIDD 189  
Db 1073 ----QSVDLL-KRFE 1083

RESULT 14

T40242  
DNA polymerase delta large chain - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C:Accession: T40242  
R:Borzym, K.; Beck, A.; Reinhardt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, October 1999  
A:Reference number: Z21916  
A:Accession: T40242  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1086 <BOR>  
A:Cross-references: EMBL:AL121815; PIDN:CAB58156.1; GSPDB:GN00067; SPDB:SPBC336.04  
A:Experimental source: strain 972h-; cosmid c336  
C:Genetics:  
A:Gene: SPDB:SPBC336.04  
A:Map position: 2  
A:Introns: 77/1  
C:Superfamily: herpesvirus DNA-directed DNA polymerase

Query Match 7.9%; Score 80; DB 2; Length 1086;  
Best Local Similarity 27.9%; Pred. No. 27;  
Matches 38; Conservative 20; Mismatches 36; Indels 42; Gaps 11;

QY 71 IKMSTPAVKNMPSFANLPLRFRGNVQERSAGATANLPLRSGR-----NMEVSLVR 122  
Db 973 ISMAAPSVGGI-----MKFA--VKVETCLGCKA--PIKKGKTALENCNLSAEYQ 1020  
QY 123 R----VPNLPQRFGRTRTTAKSVCRMLSDLCQGSMHSP--CAN---DLFYSMTQHOBIQN 173

Db 1021 ROVAQVNDLEVRFARLWTQ---CQR-----COGSMHQDVICTSRDCPIFYMRIAEHKKLQ- 1072  
QY 174 PDQKQSRRLFLFKIDD 189  
Db 1073 ----QSVDLL-KRFE 1083

RESULT 15

AB0168  
probable cell division protein ftsk [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AB0168  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentic  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Douga  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Ba  
nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AB0168  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1305 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC90205.1; PID:q15979425; GSPDB:GN00175  
C:Genetics:  
A:Gene: ftsk

Query Match 7.9%; Score 80; DB 2; Length 1305;  
Best Local Similarity 25.4%; Pred. No. 34;  
Matches 29; Conservative 24; Mismatches 41; Indels 20; Gaps 6;

QY 65 WGPKNVKNMSTPAVKNMPSFANLPLRFRGNVQERSAGATANLPLRS 112  
Db 1072 WKPSDSMDISPPMLVKLPYIVVWVDEFAFLMTVGVKVEELIARLAQKARAAGIHLVLAT 1131  
QY 113 GR---NMEVSLVRVFNLPQRFGRTRTTAKSVCRMLSDLCQGSMHSPCA-NDLFY 162  
Db 1132 QRPSVDVITGLIK--ANIPTRIAFTVSSKIDSRITLD--QGAESLLGMDMLY 1181

Search completed: March 26, 2003, 10:16:54  
Job time : 48 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 10:08:54 ; Search time 26 Seconds  
(without alignments)  
312.668 Million cell updates/sec

Title: US-09-831-758A-8  
Perfect score: 1018  
Sequence: 1 MEITSSKLFILLTATSSLL.....KQSRRLFFKKIDDAELKQEK 196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	97	9.5	931	1	POOL_ECOLI	P31828	escherichia
2	86.5	8.5	1211	1	EX5B_HAEIN	P45157	haemophilus
3	84.5	8.3	685	1	SNWA_DICDI	P54705	dictyostell
4	84	8.3	629	1	DBP6_YEAST	P53734	saccharomyc
5	81.5	8.0	1358	1	SIR4_YEAST	P11978	saccharomyc
6	80	7.9	1086	1	DPOD_SCHPO	P30316	schizosacch
7	79.5	7.8	449	1	PIV2_ADE02	P03272	human adeno
8	79.5	7.8	449	1	PIV2_ADE05	P03271	human adeno
9	79	7.8	609	1	YLO1_YEAST	P03273	human adeno
10	77	7.6	448	1	PIV2_ADE07	P48752	human adeno
11	76.5	7.5	446	1	PIV2_ADE40	P13646	homo sapien
12	76.5	7.5	458	1	K1CM_HUMAN	Q99661	homo sapien
13	75	7.4	725	1	MCAC_HUMAN	P93846	sorghum bic
14	74.5	7.3	492	1	CP51_SORBI	P02566	caenorhabdi
15	74.5	7.3	1966	1	MYSB_CAEEL	Q92185	mus muscucu
16	74	7.3	662	1	PDIL_MOUSE	Q59105	alcaligenes
17	73.5	7.2	643	1	NOS2_ALCEU	P35601	mus muscucu
18	73.5	7.2	1131	1	AC15_MOUSE	Q10172	schizosacch
19	73.5	7.2	1794	1	YAVI_SCHPO	P93596	tritricum ae
20	73	7.2	453	1	CP51_WHEAT	Q9SVN5	arabidopsis
21	73	7.2	797	1	SYM_ARATH	Q60132	citrobacter
22	72.5	7.1	330	1	T2C9_CITFR	P08730	mus muscucu
23	72.5	7.1	437	1	K1CM_MOUSE	P57211	buchnera ap
24	72.5	7.1	440	1	MESJ_BUCAI	Q10326	schizosacch
25	72.5	7.1	720	1	YD71_SCHPO	P23291	saccharomyc
26	72	7.1	538	1	K111_YEAST	P06400	homo sapien
27	72	7.1	928	1	RB_HUMAN	P14240	lymphocytic
28	72	7.1	2210	1	RRPO_LYCVA	P39705	saccharomyc
29	71.5	7.0	539	1	YAH2_YEAST	Q20224	caenorhabdi
30	71	7.0	161	1	FAB1_CAEEL	P81288	bacillus st
31	71	7.0	198	1	RS4_BACST	P50139	pichia guil
32	71	7.0	344	1	GCH2_PICGU	Q10432	schizosacch
33	71	7.0	735	1	YDD7_SCHPO		

RESULT 1  
POOL\_ECOLI  
ID POOL\_ECOLI STANDARD: PRT: 931 AA.  
AC P31828; P31829; P76132; P78158;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable zinc protease pqqL (EC 3.4.99.-).  
GN PQQ OR B1494.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID:562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RA Turlin E., Gasser F., Biville F.;  
RT "Sequence and functional analysis of an Escherichia coli DNA fragment  
able to complement pqqE and pqqF from Methylobacterium organophilum.";  
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC MEDLINE-97426617; PubMed-9278503;  
RX Blatner F.R. Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-97251357; PubMed-9097039;  
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
Sampel G., Seki Y., Sivasubraman S., Tagami H., Takeda J., Taki T.,  
Takenoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377(1996).  
CC -I- SIMILARITY: STRONG, TO H.INFLUENZAE H11368.  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.  
CC -I- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
FRAMESHIFT IN POSITION 651 THAT PRODUCES TWO SEPARATE ORFS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC

000507 h probable  
Q01036 herpesvirus  
O62954 picea abies  
O62940 pinus thunb  
Q42698 catharanthu  
P49997 azotobacter  
O94915 drosophila  
O14186 schizosacch  
P75157 mycoplasma  
P23645 drosophila  
P22338 louping ill  
O54827 mus musculu

ALIGNMENTS

```

DR EMBL; X71917; CAA50734.1; ALT_FRAME.
DR EMBL; X71917; CAA50735.1; ALT_FRAME.
DR EMBL; AE00246; AAC74567.1; -.
DR EMBL; D90791; BAA15164.1; ALT_INIT.
DR MEROPS; M16.UPB; -.
DR EcoGene; EG11744; pqqL.
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT METAL 80 80 ZINC (BY SIMILARITY).
FT ACT_SITE 83 83 BY SIMILARITY.
FT METAL 84 84 ZINC (BY SIMILARITY).
FT METAL 160 160 ZINC (BY SIMILARITY).
FT CONFLICT 360 360 A -> V (IN REF. 1).
FT CONFLICT 867 867 L -> V (IN REF. 1).
FT SEQUENCE 931 AA; 104656 MW; 94A340CA83DB6D1E CRC64;

Query Match 9.5%; Score 97; DB 1; Length 931;
Best Local Similarity 27.3%; Pred. No. 0.15;
Matches 38; Conservative 22; Mismatches 55; Indels 24; Gaps

QY 1 MEIISKILFLTLATSLTSLTINIFCA-----DELVSNLHSEKNYKYPGRGPKGER 55
   |||| | |||| | || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 1 MEIIMHNLCFLTLVATLLPLGRLTAAALPDQEXLITGLDNGLGRYMIY--PHAHPKQDV 58
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 56 SLNF-----EELKDMPGKNVIK-----MSPAVNMKPHSEANLPLRFRGNVQVEIS 101
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 59 NLWLQHTGSLQEDNELGVAHFVEHMFNGTKTPGKNKVIETESMGLRFGRDVNAVTS 118
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 102 AGAT---ANLPIRSGNME 117
   | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 119 YDETYYQVSLPTTQKONLQ 137
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 2
ID EX5B_HAEIN STANDARD; PRT; 1211 AA.
AC P45157;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
OS RECB OR H11321.
GN Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7342800;
RA Frieschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Krelschmann A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Shurin G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.B., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.J., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC phosphogluconolactides.
CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS: RECB, RECC AND RECD

```

```
CC EMBL; U43887; A840497.1; -
DR DictyDB; DD00074; SNA.
DR InterPro; IPR004015; SKIP_SNW.
DR Pfam; PF02731; SKIP_SNW; 1.
FT DOMAIN 31 41 POLY-GLN.
FT DOMAIN 194 360 SNW.
FT DOMAIN 245 253 PRO-RICH.
FT DOMAIN 409 415 POLY-ASP.
FT DOMAIN 539 616 SH2-LIKE DOMAIN.
SQ SEQUENCE 685 AA; 78529 MW; 1DC8521E9997A583 CRC64;

Query Match 8.3%; Score 84.5; DB 1; Length 685;
Best Local Similarity 23.8%; Pred. No. 1.7; Indels 37; Gaps 10;
Matches 44; Conservative 29; Mismatches 75;

QY 34 NLASKENYKDYSPRGYKGER--SLNFEELKDWGPKNYKRMSTPAVNKMPHSFANL--- 88
Db 13 NVYSNEEDPLQPKPKPOOQOQQOQOELND-RPKKVIPTYGNKGYLPRKNIEDFGDG 71
QY 89 -----PURFGNVOERSAGATANLPLRSGRNMEVSLVRRVNPPLPQRFGETTTA 137
Db 72 GAPPEIHVQYPLDMGRKKGKSSNTSNTNM---NGGGTTTIV-----PVSVDSTGRV 122
QY 138 KSVCRMSLDLCOGSMHSPCANLDFYSMTQHQEIQNPDKQSRLL-----FKKIDDAE 191
Db 123 KHEA-ILGE--KGLSHSQ-YKDLIPKQHTHE-ELQRPDDELOETLDRTKNALEKIVNGK 177
QY 192 LKQEK 196
Db 178 IKSSK 182

RESULT 4
DBP6_YEAST STANDARD; PRT; 629 AA.
AC DBP6_YEAST
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE ATP-dependent RNA helicase DBP6 (DEAD-box protein 6).
GN DBP6 OR YNR038W OR N3302.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=98187606; PubMed=9528757;
RA Kressler D., de la Cruz J., Rojo M., Linder P.;
RT "Dbp6p is an essential putative ATP-dependent RNA helicase required for 60S-ribosomal-subunit assembly in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 18:1855-1865(1998).
CC -1- FUNCTION: ATP-BINDING RNA HELICASE INVOLVED IN THE BIOGENESIS OF 60S RIBOSOMAL SUBUNITS AND IS REQUIRED FOR THE NORMAL FORMATION OF 25S AND 5.8S RRNAS.
CC -1- SUBUNIT: Interacts with DBP9.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; 271653; CAA96318.1; -
```

```
DR SGD: S0005321; DBP6.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000629; DEAD_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDc; 1.
DR SMART; SM00490; HELICC; 1.
DR PROSITE; PS00039; DEAD-ATP_HELICASE; 1.
KW Hydrolase; Helicase; ATP-binding; RNA-binding; Nuclear protein;
KW rRNA processing; Ribosome biogenesis
FT NP_BIND 234 241 ATP (POTENTIAL).
FT SITE 341 344 DEAD BOX.
SQ SEQUENCE 629 AA; 70361 MW; 64A04A2D88F44072 CRC64;

Query Match 8.3%; Score 84; DB 1; Length 629;
Best Local Similarity 21.7%; Pred. No. 1.7; Indels 44; Gaps 6;
Matches 34; Conservative 28; Mismatches 75;

QY 52 KGERSLNFEELKDWGPKNYKRMSTPAVNKMPHSFANLPLRFGNVOERSAGATANLPLR 111
Db 342 EADRLN-QSFGWCPCPLMSHLKTDKLDTP-----GNVIMFESATLTNTTEKL 390
QY 112 SGRNM-----EYSLVRRVNPPLPQRFG-RTTAKSVCR---MLSDLCQGSWHSPCAN 159
Db 391 NGLNLYKPKLFLKQTDKLYQLPKNLMEFNINPTAKSVYKPLILLYSICQFMAHSPIA-- 448
QY 160 LYSMTQHQEIQNPDKQSRLLFKKIDDAELKQEK 196
Db 449 -----AKILIFVKSNESSIRLSK 466

RESULT 5
SIR4_YEAST STANDARD; PRT; 1358 AA.
AC SIR4_YEAST
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Regulatory protein SIR4 (Silent information regulator 4).
GN SIR4 OR STE9 OR ASD1 OR UTH2 OR YDR227W OR YD9934.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88142836; PubMed=3325825;
RA Marshall M., Mahoney D., Rose A., Hicks J.B., Broach J.R.;
RT "Functional domains of SIR4, a gene required for position effect regulation in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 7:4441-4452(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95192063; PubMed=7895847;
RA Davies C.J., Hutchison C.A. III;
RT "Insertion site specificity of the transposon Tn3.";
RL Nucleic Acids Res. 23:507-514(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE PROTEINS SIR1 THROUGH SIR4 CONTROL THE EXPRESSION OF MATING TYPE GENES RESIDENT AT LOCI OF EITHER END OF CHROMOSOME III. SIR3 AND SIR4 ASSOCIATE WITH THE C-TERMINUS OF RAPI TO FORM A DNA-BINDING COMPLEX THAT INITIATES THE REPRESSION AT THE HM LOCI AND TELOMERES.
CC -1- SUBUNIT: INTERACTS WITH RAPI C-TERMINUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; 271653; CAA96318.1; -
```

Best Local Similarity	27.9%;	pred. No. 8.6;			
Matches	38;	Conservative	20;	Mismatches	36;
				Indels	42;
				Gaps	11;

```

DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Maturation protein (Protein IVA2).
GN PIVA2.
OS Human adenovirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28285;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=82211179; PubMed=7343420;
RA van Bevern C.P., Maat J., Dekker B.M.M., van Ormondt H.;
RX "The nucleotide sequence of the gene for protein IVa2 and of the 5'
RT leader segment of the major late mRNAs of adenovirus type 5.";
RL Gene 16:179-189(1981).
RN [2]
COMPLETE GENOME.
RP MEDLINE=92087470; PubMed=1727603;
RA Chroboczek J., Bieber F., Jacrot B.;
RX "The sequence of the genome of adenovirus type 5 and its comparison
RT with the genome of adenovirus type 2.";
RL Virology 186:280-285(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M73260; ; NOT_ANNOTATED_CDS.
DR DR
EMBL; X02996; CAB40667.1; .
PIR; A03842; QADAA5.
DR DR
InterPro; IPR003389; Adeno_IVa2.
DR DR
Pfam; PF02456; Adeno_IVa2; 1.
DR KW
ATP-binding.
FT NP_BIND 171 178 ATP (POTENTIAL).
SQ SEQUENCE 449 AA; 50887 MW; 20AD30752DCA96C2 CRC64;
Query Match 7.8%; Score 79.5; DB 1; Length 449;
Best Local Similarity 25.0%; Pred. No. 3.1; Indels 23; Gaps 8;
Matches 40; Conservative 27; Mismatches 70;
QY 28 DELVMNLHNSKENDKYSEPRGY-----PKGERSLNFEELKD--WGPKNVTK----MSTP 76
| : : : | | | | : : : | : : | : : | :
DB 249 DDLIL-----EHNYD-VSDPNIFAQAARGPIAIDMECKNGLGGHKGVSKFFHAFP SK 302
| : : : | | | | : : : | : : | : : | :
QY 77 AVNKMPHSFANLPFRGNVOEERS-ACATANLPURSGNNNEVSLVRVPNLPQRFGRTT 135
| : : | : : | : : | : : | : : | : : | : : | :
DB 303 LHDKEPKCTGTVLVLVNMRPRDMAGNIANKIKTS--KMHLISPRMHPSQLNFVNTY 360
| : : | : : | : : | : : | : : | : : | : : | :
QY 136 T---AKSVCRMLSDLCOGSMHSPCANLDFYSMTQCORIEQ 172
| : : | : : | : : | : : | : : | : : | : : | :
DB 361 TKGLPLAISLLKDI FRHHAQSCYDWIYYNTTPQEALQ 400
| : : | : : | : : | : : | : : | : : | : : | :
RESULT 9
YL01_YEAST
ID YL01_YEAST STANDARD; PRT: 609 AA.
AC Q06053; P39519;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 69.8 kDa protein in BDF1-SFP1 intergenic region.
GN YLR041C OR L8084.19.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,

```

RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifkin L., Riles L., Taich A., Trevaskis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.,  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RN SEQUENCE OF 504-609 FROM N.A.  
 RP STRAIN-S288c;  
 RC MEDLINE-9511623; PubMed-7816623;  
 RX Lygerou Z., Conesa C., Lesage P., Swanson R.N., Ruet A., Carlson M.,  
 RA Sentenac A., Seraphin B.,  
 RT "The yeast Bp1 gene encodes a transcription factor involved in the  
 expression of a broad class of genes including snRNAs.";  
 RL Nucleic Acids Res. 22:5332-5340(1994).  
 CC - SIMILARITY: BELONGS TO THE UPF0034 (NIF3/SMML) FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U19729; AAB82358.1; -  
 DR EMBL; Z18944; CAA79376.1; -  
 DR SGD; S0004393; YLR401C.  
 DR InterPro; IPR001269; UPF0034.  
 DR Pfam; PF00642; zf-CCCH; 1.  
 DR Pfam; PF01207; UPF0034; 1.  
 DR PROSITE; PS01136; UPF0034; 1.  
 KW Hypothetical protein.  
 FT CONFLICT 602 602 W -> M (IN REF. 2).  
 FT SEQUENCE 609 AA; 69817 MW; 948239C5B0F75909 CRC64;  
 SQ  
 Query Match 7.8%; Score 79; DB 1; Length 609;  
 Best Local Similarity 21.7%; Pred. No. 5.1;  
 Matches 35; Conservative 25; Mismatches 69; Indels 32; Gaps 4;  
 QY 29 ELVMSNLHSEKNDYSEPRGYPKGRSLNFELKDWGPKNVKIMSTPAVNKMPHSFANL 88  
 DB 261 ELSEHRMKQREVLKDYKDRYFAQEKPLDLYHKKIVSPLTTV-----GNL 306  
 QY 89 PLR-----FCRNVOEERSAGATANPLRSGRNMEVSLVRRVNPVLPORFG-----R 133  
 DB 307 PYRLMRKLGADTYISEMALA---VPLIOGTNSEWALPKAHTSEFFGFGVQVACSKAWQA 363  
 QY 134 TTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTCHQHEIQNP 174  
 DB 364 AKAAEALANSVSEIENLNSGCPIDLLYRQSGSALLDNP 404  
 RESULT 10  
 PIV2\_ADE07  
 ID PIV2\_ADE07 STANDARD; PRT; 448 AA.  
 AC P03273; Q64833;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Maturation protein (Protein IVA2).  
 GN PIV2.  
 OS Human adenovirus type 7.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=10519;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-83054637; PubMed-6292051;  
 RX Engler J.A., van Bree M.P.;  
 RA "The nucleotide sequence of the gene encoding protein Iva2 in human  
 RT adenovirus type 7.";  
 RT Gene 19:71-80(1982).  
 RL

[2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-Gomen;  
 RA van Ormondt H.;  
 RL Submitted (JUN-1985) to the EMBL/GenBank/DBJ databases.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X03000; CAA26766.1; -  
 DR PIR; A03843; Q4ADA7.  
 DR InterPro; IPR003389; Adeno\_IVa2.  
 DR Pfam; PF02456; Adeno\_IVa2; 1.  
 KW ATP-binding.  
 FT NP\_BIND 171 178 ATP (POTENTIAL).  
 FT CONFLICT 306 306 K -> R (IN REF. 2).  
 FT SEQUENCE 448 AA; 50678 MW; 01F17E29320848BC CRC64;  
 SQ  
 Query Match 7.6%; Score 77; DB 1; Length 448;  
 Best Local Similarity 23.4%; Pred. No. 5.5;  
 Matches 39; Conservative 26; Mismatches 68; Indels 34; Gaps 8;  
 QY 29 ELVMSNLHSEKNDYSEPRGYPKGRSLNFELKDWGPKNVKIMST-----PAVNKM 81  
 DB 245 KWAYDDLTOEHNYD-VSDPRNV-----FARAAAHGPIAIIIDECMENLGCHKGVSFK 295  
 QY 82 PHSF-ANLPLRFR-----NVQERSAGAT-ANPLRSGRNMEVSLVRRVNPVLP 128  
 DB 296 FHAFPSKLHDKPKCKGTGYTVLVVHLNHNPRDLGGNIANKIOS--KMHIISPRMHPSQL 353  
 QY 129 QRFGRITTT---AKSVCRMLSDLCOGSMHSPCANDLFYSMTCHQHEIQ 172  
 DB 354 NRPVNTYTTKGLPVAISLLKLDIVOHHLRPCYDVIYNTTPEALQ 400  
 RESULT 11  
 PIV2\_ADE40  
 ID PIV2\_ADE40 STANDARD; PRT; 446 AA.  
 AC P48752;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Maturation protein (Protein IVA2).  
 GN PIV2.  
 OS Human adenovirus type 40.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=28284;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Dugan;  
 RX MEDLINE-94087748; PubMed-8263936;  
 RA Davison A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;  
 RT "The DNA sequence of adenovirus type 40.";  
 RL J. Mol. Biol. 234:1308-1316(1993).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L19443; AAC13952.1; -  
 DR InterPro; IPR003389; Adeno\_IVa2.  
 DR Pfam; PF02456; Adeno\_IVa2; 1.  
 KW ATP-binding.  
 FT NP\_BIND 170 177 ATP (POTENTIAL).  
 FT

```

KERATIN 13 IS GENERALLY ASSOCIATED WITH KERATIN 4.
-!- DISEASE: Defects in KRT13 are a cause of white sponge nevus (WSN). WSN is a rare autosomal dominant disorder which predominantly affects noncornified stratified squamous epithelia. Clinically, it is characterized by the presence of soft, white, and spongy plaques in the oral mucosa. The characteristic histopathologic features are epithelial thickening, parakeratosis, and vacuolization of the suprabasal layer of oral epithelial keratinocytes. Less frequently the mucous membranes of the nose, esophagus, genitalia and rectum are involved.
-!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND MICROFILAMENTARIL KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].
-!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licens@isb-sib.ch).

EMBL; X14640; CAA32786.1; -
EMBL; X52426; CAA36673.1; -
EMBL; AF049259; AAC35754.1; -
EMBL; BC002661; AAH02661.1; -
PIR; S06086; KRH03.
PIR; A33403; A33403.
GlycoSuiteDB; P13646; -.
Genew; HGNC:6415; KRT13.
MIM; 148065; -.
MIM; 193900; -.
InterPro; IPRO01664; IF.
InterPro; IPRO02957; Keratin_I.
Pfam; PF00038; filament; 1.
PRINTS; PR01248; TYPE1KERATIN.
PROSITE; PS00226; IF_1.
Intermediate filament; Coiled coil; Keratin; Disease mutation.
DOMAIN 1 103 HEAD.
FT DOMAIN 104 412 ROD.
FT DOMAIN 413 458 TAIL.
FT DOMAIN 104 139 COIL 1A.
FT DOMAIN 140 158 LINKER 1.
FT DOMAIN 159 250 COIL 1B.
FT DOMAIN 251 273 LINKER 12.
FT DOMAIN 274 412 COIL 2.
FT VARIANT 119 119 L -> P (IN WSN).
FT /FTid=VAR_003836.
CONFLICT 58 58 G -> D (IN REF. 1 AND 2).
CONFLICT 416 457 MIGFPPSAGSVSPRSSTVTTSSAVTTTSNASGRRTSDVR
SQ R -> KRQP (IN REF. 3 AND 4).
SEQUENCE 458 AA; 49586 MW; 9E033A247CD2B106 CRC64;

Query Match 7.5%; Score 76.5; DB 1; Length 458;
Best Local Similarity 23.4%; Pred. No. 6.3;
Matches 41; Conservative 28; Mismatches 71; Indels 35; Gaps 8;

QY 28 DELVMNLHSKENYDKYSEPRGYPKGRSLNFELKDW-----GPKNVYIKMTSAVN--- 79
||| : | : : | : | : ||| : | : | : | : | :
Db 216 DELTLSTKLEMQIESLNHEELAYMKKNHE---EEMKFESNQVVGVNVEDATPGIDLTR 272
| : | : | : | : | : | : | : | : | : | : | :
QY 80 ---KMPSHFANLPURFGNRVOE-----ERSAGATANLPURSGRNMEVSLVRV---- 124
: | : | : | : | : | : | : | : | : | : | :
Db 273 VLAEMREQYEAMERNRDAPWEPFAKSALINKVEVNTAMIQTSKTEITRLTQGLE 332
| : | : | : | : | : | : | : | : | : | : | :
QY .125 ----PNLPQREG-RTTYAKSVCRMLSDL--COGSWHSPCA--NDLFYSMTCOHOE 170
| : | : | : | : | : | : | : | : | : | : | :
Db 333 IEQLQSLSMRKAGLENTVAETECRYALQLOQQIOLSIIEALQSELURSEMEEQNQE 387
| : | : | : | : | : | : | : | : | : | : | :

RESULT 13
MCAN_HUMAN

```



DB 379 PKGHVATSPFANLPHIYKPDSDYDFRCPGREDKAAGAFYSISFGGRH 432

RESULT 15

MYSB\_CAEEL STANDARD; PROT; 1966 AA.

AC P02566;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE MYOSIN heavy chain B (MHC B).

GN UNC-54 OR MYO-4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_Taxid=6239;

RN (1)

RP SEQUENCE FROM N.A.

RA MEDLINE=83273600; PubMed=6576334;

RA Karn J., Brenner S., Barnett L.;

RT "protein structural domains in the Caenorhabditis elegans unc-54

RT myosin heavy chain gene are not separated by introns.;"

RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).

RN (2)

RP SEQUENCE OF 850-1966 FROM N.A.

RA MEDLINE=82272395; PubMed=7202124;

RA McLachlan A.D., Karn J.;

RT "periodic charge distributions in the myosin rod amino acid sequence

RT match cross-bridge spacings in muscle.;"

RL Nature 299:226-231(1982).

RN (3)

RP SEQUENCE OF 1876-1966 FROM N.A.

RA MEDLINE=83232892; PubMed=6571695;

RA Willis N., Gesteland R.F., Karn J., Barnett L., Bolten S.,

RA Waterston R.H.;

RT "The genes sup-7 X and sup-5 III of C. elegans suppress amber

RT nonsense mutations via altered transfer RNA.;"

RL Cell 33:575-583(1983).

CC -1- FUNCTION: MUSCLE CONTRACTION.

CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY

CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT

CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE

CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED

CC SUBFRAGMENT (S2).

CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN

CC C. ELEGANS.

CC -1- WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.

CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL: J01050; AAA28124.1; -

CC EMBL: V01494; CAA24738.1; -

CC PIR: A02992; MKRW.

CC HSSP: P08799; IMNO.

CC InterPro: IPR004009; Myosin\_N.

CC InterPro: IPR002928; Myosin\_tail.

CC InterPro: IPR001609; myosin\_head.

DR Pfam: PF00063; myosin\_head; 1.

DR Pfam: PF01576; Myosin\_tail; 1.

DR Pfam: PF02736; Myosin\_N; 1.

DR PRINTS: PR00193; MYOSINHEAVY.

DR PRODOM: PD000355; myosin\_head; 1.

DR SMART: SM00242; MYSC; 1.

KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;

KW ATP-binding; Methylation; Alkylation; Multigene family.

FT DOMAIN 1 850 MYOSIN HEAD-LIKE.

FT DOMAIN 851 1164 COILED COIL (POTENTIAL).

FT DOMAIN 1165 1176 ALPHA-HELICAL TAILPIECE (S2).

FT DOMAIN 1165 1966 HINGE.

FT NP\_BIND 177 184 LIGHT MEROMYOSIN (LMW).

FT DOMAIN 665 687 ATP (BY SIMILARITY).

FT DOMAIN 769 783 ACTIN-BINDING.

FT MOD\_RES 128 128 METHYLATION (TRI-) (POTENTIAL).

FT MOD\_RES 705 705 ALKYLATION (SH-1).

FT MOD\_RES 715 715 ALKYLATION (SH-2).

FT CONFLICT 1337 1337 E -> R (IN REF. 2).

FT CONFLICT 1880 1880 I -> L (IN REF. 2).

SQ SEQUENCE 1966 AA; 225125 MW; B66F0BB2FE27B67F CRC64;

Query Match 7.3%; Score 74.5; DB 1; Length 1966;

Best Local Similarity 19.2%; Pred. No. 63;

Matches 47; Conservative 46; Mismatches 101; Indels 51; Gaps 10;

QY 1 METISSKFLITLATSSLL--TSNIFCADELVMNLHSHKENYDKYSEPRGYPKGRSLN 58

DB 523 IELIEKPLGIISMLDEECIVPKATDLTLASKLVQHLGKHPNEKPKPKG-KQGEARFA 581

QY 59 FEELK-----DWGPKVYKMTSPAVNKMPSFAN---LPLRFGNVOEERSA----- 102

DB 582 MRHYACTVRYNCLNWLKKNKNDPLNDTVVSAMKQSGKNDLVEIWQDYTTQEEAAKAKEG 641

QY 103 GATANLPLSGRNNEVSLVR-----VPNLPRFGRTTAA---- 137

DB 642 GGGGKKKKGSGSPMTVMCLYRESLNNLMTMLNKTHTPHFIRCIIPNEKKQSGMIDALVLN 701

QY 138 KSVCRMLSD---LCQSGMHSFPCANDLF---YSMTQCQHOEIQNPDKQ-SRRLLPKKIDDA 190

DB 702 QLTGCVLEIGIRICRKGFPNRTLHPDFVQRYAILAAKEAKSDDDKKCAEAIMSKLVNDG 761

QY 191 ELKQE 195

DB 762 SLSEE 766

Search completed: March 26, 2003, 10:14:30

Job time : 30 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: March 26, 2003, 10:16:10 ; Search time 16 Seconds  
(without alignments)  
719.365 Million cell updates/sec

Title: US-09-831-758A-8  
Perfect score: 1018  
Sequence: 1 MEIISKLFILLATSSLL.....KQSRLLFKKIDDAELKQEK 196

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues  
Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pap.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1014	99.6	196	US-09-995-793A-29	Sequence 29, Appl
2	82.5	8.1	4019	US-09-854-133-425	Sequence 425, App
3	82.5	8.1	4019	US-09-738-973-425	Sequence 425, App
4	81	8.0	1362	US-09-815-242-14009	Sequence 14009, A
5	75.5	7.4	869	US-10-016-283-33	Sequence 33, Appl
6	75.5	7.4	869	US-09-817-487A-2	Sequence 2, Appli
7	74.5	7.3	1170	US-10-135-687-2	Sequence 20, Appl
8	73	7.2	280	US-09-796-138-20	Sequence 20, Appl
9	73	7.2	280	US-09-909-903-20	Sequence 20, Appl
10	73	7.2	648	US-09-815-242-10504	Sequence 10504, A
11	72.5	7.1	348	US-09-840-787-10	Sequence 10, Appl
12	72	7.1	781	US-09-469-522-37	Sequence 37, Appl
13	72	7.1	797	US-09-469-522-45	Sequence 45, Appl
14	72	7.1	816	US-09-469-522-4	Sequence 4, Appli
15	72	7.1	832	US-09-469-522-35	Sequence 35, Appl
16	72	7.1	851	US-09-469-522-33	Sequence 33, Appl
17	72	7.1	851	US-09-469-522-39	Sequence 39, Appl
18	72	7.1	859	US-09-469-522-43	Sequence 43, Appl
19	72	7.1	869	US-09-469-522-47	Sequence 47, Appl

20	72	7.1	871	10	US-09-469-522-49	Sequence 49, Appl
21	72	7.1	874	10	US-09-469-522-31	Sequence 31, Appl
22	72	7.1	895	10	US-09-469-522-29	Sequence 29, Appl
23	72	7.1	897	10	US-09-469-522-41	Sequence 41, Appl
24	72	7.1	928	10	US-09-758-007-3	Sequence 3, Appli
25	72	7.1	928	10	US-09-860-211-8	Sequence 8, Appli
26	72	7.1	928	10	US-09-469-522-2	Sequence 2, Appli
27	72	7.1	928	10	US-09-469-522-51	Sequence 51, Appl
28	71	7.0	147	9	US-09-738-626-4146	Sequence 4146, Ap
29	71	7.0	377	10	US-09-934-778-2	Sequence 2, Appli
30	70.5	6.9	273	10	US-09-815-242-11993	Sequence 11993, A
31	70	6.9	1508	9	US-10-024-623-35	Sequence 35, Appl
32	70	6.9	1508	12	US-10-002-769-15	Sequence 15, Appl
33	69.5	6.8	747	9	US-09-978-295A-459	Sequence 459, App
34	69.5	6.8	747	9	US-09-938-418-9	Sequence 9, Appli
35	69.5	6.8	747	9	US-09-978-697-459	Sequence 459, App
36	69.5	6.8	747	9	US-09-978-132A-459	Sequence 459, App
37	69.5	6.8	747	9	US-09-939-832A-459	Sequence 459, App
38	69.5	6.8	747	9	US-09-978-189-459	Sequence 459, App
39	69.5	6.8	747	9	US-10-028-072-426	Sequence 426, App
40	69.5	6.8	747	9	US-10-121-049-426	Sequence 426, App
41	69.5	6.8	747	9	US-10-123-904-426	Sequence 426, App
42	69.5	6.8	747	9	US-10-140-470-426	Sequence 426, App
43	69.5	6.8	747	9	US-10-175-746-426	Sequence 426, App
44	69.5	6.8	747	9	US-10-176-918-426	Sequence 426, App
45	69.5	6.8	747	9	US-10-176-921-426	Sequence 426, App

ALIGNMENTS

RESULT 1  
US-09-995-793A-29  
; Sequence 29, Application US/09995793A  
; Publication No. US20030054446A1  
; GENERAL INFORMATION:  
; APPLICANT: Weber, Bernard H.F.  
; APPLICANT: Stoehr, Heidi  
; TITLE OF INVENTION: No. US20030054446A1 retina-specific human proteins C7ori  
; FILE REFERENCE: 033488-001  
; CURRENT APPLICATION NUMBER: US/09/995,793A  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: 60/253,751  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 29  
; TYPE: PRT  
; LENGTH: 196  
; ORGANISM: Homo sapiens  
US-09-995-793A-29

Query Match 99.6%; Score 1014; DB 9; Length 196;  
Best Local Similarity 99.5%; Pred. No. 7.8e-103;  
Matches 195; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MEIISKLFILLATSSLLTSNIFC	ADLVMSNLHSENYDKYSEPRGYPKGRSLNFE	60
DB	1	MEIISKLFILLATSSLLTSNIFC	ADLVMSNLHSENYDKYSEPRGYPKGRSLNFE	60
QY	61	ELKDWGPKNYKSTPAVNKMPHSFANL	PLRFRGNVOEERSAGATANLPLRSGRNMEVSL	120
DB	61	ELKDWGPKNYKSTPAVNKMPHSFANL	PLRFRGNVOEERSAGATANLPLRSGRNMEVSL	120
QY	121	VRRVNPQRFGRTRTTAKSVCRMLSD	LCQSGSMHSPCANDLFYSMTQHOETQNPDKOSR	180
DB	121	VRRVNPQRFGRTRTTAKSVCRMLSD	LCQSGSMHSPCANDLFYSMTQHOETQNPDKOSR	180
QY	181	RLLFKKIDDAELKQEK	196	
DB	181	RLLFKKIDDAELKQEK	196	

RESULT 2  
US-09-854-133-425  
; Sequence 425, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854.133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 425  
; LENGTH: 4019  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-425

Query Match 8.1%; Score 82.5; DB 9; Length 4019;  
Best Local Similarity 20.6%; Pred. No. 21;  
Matches 34; Conservative 27; Mismatches 61; Indels 43; Gaps 5;

QY 28 DELVMSNLHSHKENVKDYSEP-----RGYPKGRSLNFEELKDWGPKNVIKM 73  
DB 3733 EDLVLSDISPKGVWDKILEPVACVRRKKSEMLQLFPAYLKGE-----DLFGL 3778  
QY 74 STPAVNMKPHSF-----ANLPLFRGNVQVEERSAGATANLPLRSGRNMEVSLVRRVPNL 127  
DB 3779 TVSAVARIAESLPGVEACENYFRYGRNPLMELPLAVNPTGCARSEPKMSAHVVR-----3833  
QY 128 PORFGRRTTTAKSVCRMSLDLCQSMHSPCANDLFYSMTQHQEIQ 172  
DB 3834 PHTLNSSTSKS-----FQSTVTGELNAPYSKQFVHSSKSSQYRKMK 3874

RESULT 3  
US-09-738-973-425  
; Sequence 425, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738.973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 425  
; LENGTH: 4019  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-738-973-425

Query Match 8.1%; Score 82.5; DB 10; Length 4019;  
Best Local Similarity 20.6%; Pred. No. 21;  
Matches 34; Conservative 27; Mismatches 61; Indels 43; Gaps 5;

QY 28 DELVMSNLHSHKENVKDYSEP-----RGYPKGRSLNFEELKDWGPKNVIKM 73  
DB 3733 EDLVLSDISPKGVWDKILEPVACVRRKKSEMLQLFPAYLKGE-----DLFGL 3778  
QY 74 STPAVNMKPHSF-----ANLPLFRGNVQVEERSAGATANLPLRSGRNMEVSLVRRVPNL 127  
DB 3779 TVSAVARIAESLPGVEACENYFRYGRNPLMELPLAVNPTGCARSEPKMSAHVVR-----3833  
QY 128 PORFGRRTTTAKSVCRMSLDLCQSMHSPCANDLFYSMTQHQEIQ 172  
DB 3834 PHTLNSSTSKS-----FQSTVTGELNAPYSKQFVHSSKSSQYRKMK 3874

RESULT 4  
US-09-815-242-14009  
; Sequence 14009, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA 011A  
; CURRENT APPLICATION NUMBER: US/09/815.242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14009  
; LENGTH: 1362  
; TYPE: PRT  
; ORGANISM: Salmonella typhi  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(1362)  
; OTHER INFORMATION: Xaa - Any Amino Acid  
US-09-815-242-14009

Query Match 8.0%; Score 81; DB 10; Length 1362;  
Best Local Similarity 25.2%; Pred. No. 6.8;  
Matches 29; Conservative 24; Mismatches 42; Indels 20; Gaps 6;

QY 65 WGPKNVVKMSTPAVNMKMPH-----SFANLPLFRGNVQVE-----ERSAGATANLPLRS 112  
DB 1130 WAPGSDMDVQHPVLEKLPYIVLVDEFADLMTVCKVEELIARLAQKARAAGIHLVLAT 1189  
QY 113 GR---NMEVSLVRRVPNLQREFRRTTAKSVCRMSLDLCQSMHSPCA-NDLFYS 163  
DB 1190 QRPVSDVITGLIK--ANIPTRIAFTVSSKIDSTILD--QGAESLLGMDMLYS 1240

RESULT 5  
US-10-016-283-33

```

; Sequence 33, Application US/10016283
; Patent No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-283-33

Query Match 7.4%; Score 75.5; DB 9; Length 869;
Best Local Similarity 20.1%; Pred. No. 14;
Matches 44; Conservative 37; Mismatches 69; Indels 69; Gaps 14;

QY 1 MEIIS-KLFIILLTSLTSLTSNIFCA-----LYCRRRRKQWKKKRESAAVTLTLPSELLELRLHP 550
DB 498 ISIMSSFAIFVLLTIT-----LYCRRRRKQWKKKRESAAVTLTLPSELLELRLHP 550

QY 38 KENYOKYS---EPR-----GYPKGRSLNFEELKDWGP---KNVVKMSTPAVNKMPHS-FA 86
DB 551 NPMYQRMPLLNPKLLSLEYPRN---NIEYVRDIGEGAGRVFQARAPGL--LPYEPT 604

QY 87 NLPLRFGNVOEERSAGATANPLRSGRNMEVSLVRRV--PNLPQRFGRTTTAKSVCRML 144
DB 605 MVAVKM---LKEEASADMQADF-----QREAAALMAEFDPNPVIVKLLGVCAVGKPMCLLF 655

QY 145 SDCOG-----SMHSPCANDLFYSMTQHOEIQNP 174
DB 656 EYMAYGDLNEFLRSMSPHTVCS--LSHSDLSMRAQVSSP 692

RESULT 6
US-09-817-487A-2
; Sequence 2, Application US/09817487A
; Patent No. US20020150876A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020150876A1artis AG
; TITLE OF INVENTION: Selectable Marker Genes
; FILE REFERENCE: 4-31193A
; CURRENT APPLICATION NUMBER: US/09/817,487A
; CURRENT FILING DATE: 2002-02-14
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 869
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-817-487A-2

Query Match 7.4%; Score 75.5; DB 10; Length 869;
Best Local Similarity 20.1%; Pred. No. 14;
Matches 44; Conservative 37; Mismatches 69; Indels 69; Gaps 14;

QY 1 MEIIS-KLFIILLTSLTSLTSNIFCA-----LYCRRRRKQWKKKRESAAVTLTLPSELLELRLHP 550
DB 498 ISIMSSFAIFVLLTIT-----LYCRRRRKQWKKKRESAAVTLTLPSELLELRLHP 550

QY 38 KENYOKYS---EPR-----GYPKGRSLNFEELKDWGP---KNVVKMSTPAVNKMPHS-FA 86
DB 551 NPMYQRMPLLNPKLLSLEYPRN---NIEYVRDIGEGAGRVFQARAPGL--LPYEPT 604

QY 87 NLPLRFGNVOEERSAGATANPLRSGRNMEVSLVRRV--PNLPQRFGRTTTAKSVCRML 144
DB 605 MVAVKM---LKEEASADMQADF-----QREAAALMAEFDPNPVIVKLLGVCAVGKPMCLLF 655

QY 145 SDCOG-----SMHSPCANDLFYSMTQHOEIQNP 174
DB 656 EYMAYGDLNEFLRSMSPHTVCS--LSHSDLSMRAQVSSP 692

```

```

DB 605 MVAVKM---LKEEASADMQADF-----QREAAALMAEFDPNPVIVKLLGVCAVGKPMCLLF 655
QY 145 SDCOG-----SMHSPCANDLFYSMTQHOEIQNP 174
DB 656 EYMAYGDLNEFLRSMSPHTVCS--LSHSDLSMRAQVSSP 692

RESULT 7
US-10-135-687-2
; Sequence 2, Application US/10135687
; Patent No. US20020123120A1
; GENERAL INFORMATION:
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001068DIV
; CURRENT APPLICATION NUMBER: US/10/135,687
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 09/749,588
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-135-687-2

Query Match 7.3%; Score 74.5; DB 12; Length 1170;
Best Local Similarity 22.9%; Pred. No. 28;
Matches 49; Conservative 33; Mismatches 53; Indels 79; Gaps 12;

QY 31 VMSLNHSKENDKYSEPRGYP---KGRSLNFEELK---DWGPKNVKMKSTPAVNKMP 82
DB 237 ILSRL--SENADENYFVRSYECFOHKNHTCLVFEMLEQNLYDFLKN-----282

QY 83 HSFANLPLRFGNVOEERSAGATANPLRS-----GRNMEVSLVRRVPLNLPOR-- 130
DB 283 -KFSPLPLKTYIRPILQQ---VATMLMKLSGLIHADLKPENIMLVDPVRQ----PYRVK 334

QY 131 ---FGRTT-TAKSVCR-----MLSDLCQSGSMHSPC 156
DB 335 VDFGSASHVSKAVCSYLOSRYRAPEIILGLPCEAIDMKSLGCVIAELFLGWPLYPG 394

QY 157 ANDLFYSMTQHOEIQNP-DQKQSRLLPKKIDD 189
DB 395 ASE--YDQTPBEHELETGKSKARKYIFNCLDD 426

RESULT 8
US-09-796-138-20
; Sequence 20, Application US/09796138
; Patent No. US20020031782A1
; GENERAL INFORMATION:
; APPLICANT: Waterman, Michael R.
; APPLICANT: Bellamine, Aouatef
; APPLICANT: Podust, Larissa M.
; TITLE OF INVENTION: Mycobacterium tuberculosis cyp51 HIGH RESOLUTION STRUCTURE,
; TITLE OF INVENTION: AND
; TITLE OF INVENTION: NUCLEIC ACIDS, AND THERAPEUTIC AND SCREENING METHODS
; TITLE OF INVENTION: RELATING TO SAME
; FILE REFERENCE: Attorney Docket No. US20020031782A1 1242-17-2
; CURRENT APPLICATION NUMBER: US/09/796,138
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/345,218
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Triticum sp.

```

US-09-796-138-20

```

Query Match          7.2%; Score 73; DB 10; Length 280;
Best Local Similarity 33.3%; Pred. No. 5.6;
Matches 22; Conservative 13; Mismatches 25; Indels 6; Gaps 3;

Qy 56 SLNFELKDWG-PKNVIMSTPA-VNKMPSFANL-----PLRFGNRVQERSAGATANLP 109
   | : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | :
Db 155 SVTTREGKEFDIPKGHIVATSPAFANRLPHIFKNPDSYDPDRFAAGREEDKVGAFSYIS 214
   | : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | :
Qy 110 LRSGRN 115
   ||:
Db 215 FGGRH 220

```

RESULT 9

```

US-09-909-903-20
; Sequence 20, Application US/09909903
; Patent No. US20020052031A1
; GENERAL INFORMATION:
; APPLICANT: Waterman, Michael R.
; APPLICANT: Bellamine, Aouatef
; TITLE OF INVENTION: Mycobacterium tuberculosis CYP51 POLYPEPTIDES AND
; TITLE OF INVENTION: NUCLEIC ACIDS AND THERAPEUTIC AND SCREENING METHODS
; FILE REFERENCE: RELATING TO SAME
; FILE REFERENCE: Attorney Docket No. US20020052031A1 1242-17
; CURRENT FILING DATE: 2001-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 20
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Triticum sp.
US-09-909-903-20

```

```

Query Match          7.2%; Score 73; DB 10; Length 280;
Best Local Similarity 33.3%; Pred. No. 5.6;
Matches 22; Conservative 13; Mismatches 25; Indels 6; Gaps 3;

```

```

Qy 56 SLNFELKDWG-PKNVIMSTPA-VNKMPSFANL-----PLRFGNRVQERSAGATANLP 109
   | : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | :
Db 155 SVTTREGKEFDIPKGHIVATSPAFANRLPHIFKNPDSYDPDRFAAGREEDKVGAFSYIS 214
   | : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | :
Qy 110 LRSGRN 115
   ||:
Db 215 FGGRH 220

```

RESULT 10

```

US-09-815-242-10504
; Sequence 10504, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

```

```

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10504
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10504

```

```

Query Match          7.2%; Score 73; DB 10; Length 648;
Best Local Similarity 24.8%; Pred. No. 18;
Matches 34; Conservative 21; Mismatches 48; Indels 34; Gaps 6;

```

```

Qy 36 HSKENYDKY-----SEPRGYPKGERSLNFELKDWGPKNVIMSTPAVNKMPHSFA 86
   | : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | :
Db 241 HYKGNYSKYLELKAELASEWKAQEKEIN--KLEDFVAKNLVRASTTKRAQSRKVL 298
   | : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | :
Qy 87 NLPLRFGNRVQERSA-----GATANLPLR-----SGRNMEVSL-----VRRVPNL 127
   | : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | :
Db 299 EKMDRLDRPQGDSEKSAHFLDSEKVSQNVYLVQVDAAGYDQEHILSEPIHLDIRKEAI 358
   | : | : | | | : | : | | | : | : | | | : | : | | | : | : | | | :
Qy 128 ----PQREGRTTTAKSV 140
   ||:
Db 359 ALVGNPNGIGKSTLLKSI 375

```

RESULT 11

```

US-09-840-787-10
; Sequence 10, Application US/09840787
; Patent No. US20020058264A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Bandman, Olga
; Shah, Purvi
; Au-Young, Janice
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/840,787
; FILING DATE: 23-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: 09/518,865
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

```

```
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: MUSCNOT01
CLONE: 118160
SEQUENCE DESCRIPTION: SEQ ID NO: 10 :
US-09-840-787-10

Query Match      7.1%  Score 72.5;  DB 10;  Length 348;
Best Local Similarity 22.2%;  Pred. No. 8.6;
Matches 42;  Conservative 25;  Mismatches 81;  Indels 41;  Gaps 8:

QY 26 CADELVMNLHLSK-ENYKISEPRGYPKGERSLNFEELK-----DWGPKNVIKMST 75
Db 59 CSDKEVSLAKVLINWRLDSPGPKGEKEEREKAKKKEGLECSDWKE-----111
QY 76 PAVNKMPSFANLPLFRGNVQERSAGATANLPLRSGRNMEVSLVRRVNPQLPQRFGR 135
Db 112 -AGLSPPKKREDPKTRRDSVDKSSASSPKRPSVERSNSKSAES-PKTPSSPLTPT 169
QY 136 TAKSVCRMLSDLCQ-SMHSPC-----ANDLF--YSMTQ-----HGEION 173
Db 170 FASSMCLLAPCYLTGDSVRDKCVEMLSAALKADDDYKDYGVNCDKNASEIEDHIY 229
QY 174 PDKQSRRL 182
Db 230 TDMKYNRV 238

RESULT 12
US-09-469-522-37
; Sequence 37, Application US/09469522
; Patent No. US20020151461A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong-Ji
;           Hu, Shi-Xue
;           Zhou, Yunli
; TITLE OF INVENTION: MODIFIED RETINOBLASTOMA TUMOR SUPPRESSOR
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/469,522
; FILING DATE: 22-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/026,459
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: UTXC:506
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
```

```
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 781 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 37:
US-09-469-522-37

Query Match      7.1%  Score 72;  DB 10;  Length 781;
Best Local Similarity 21.5%;  Pred. No. 30;
Matches 47;  Conservative 23;  Mismatches 63;  Indels 86;  Gaps 8:

QY 17 SSLTSLNIF-----CADELVMS-----NLHSENKYSEPR 48
Db 327 SKLLNDNIFHMSLLACALEVMATYRSTSONLDSGTLSFPWILNVLNKAFFYK 386
QY 49 GYPKGERSLNFEELK-----DWGPKNVIKMSTPAVNKMPSFANLPLR 91
Db 387 SFKAEGNLTREMIKHLERCEHRIMESLAWLSDFLLIKQSKDREGPTDHL 445
QY 92 FGRNVQERSAGATANLPLRSGR---NMEVSLVRRVNPQLPQRFGRITTTAKSV 148
Db 446 -----NLPLQNNHTAADMILSPVS-----PKKGGSTTRVNSTANAETQ--- 484
QY 149 QGSMHSPCANDLFYSMTQHQEIQNPDQKQSRRLLEFKI 187
Db 485 -----ATSAF-----QTQKPLKSTSLSLFYKKV 507

RESULT 13
US-09-469-522-45
; Sequence 45, Application US/09469522
; Patent No. US20020151461A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong-Ji
;           Hu, Shi-Xue
;           Benedict, William F.
;           Zhou, Yunli
; TITLE OF INVENTION: MODIFIED RETINOBLASTOMA TUMOR SUPPRESSOR
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/469,522
; FILING DATE: 22-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/026,459
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: UTXC:506
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-469-522-45

Query Match          7.1%; Score 72; DB 10; Length 797;
Best Local Similarity 21.5%; Pred. No. 31;
Matches 47; Conservative 23; Mismatches 63; Indels 86; Gaps 8;

OY 17 SSLTSTNIF-----CADELVMS-----NLHSHENYDKYSEPR 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 SKLNDNFHMSLLACALEVVMATYRSTSONLDSGTDLSFPFWILNVLNKAQDFYKVE 402
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 49 GYPKGRSLNFEELK-----DWGPKNVKMTSPVANKPHSPANLPLR 91
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 SFKAEGNLTREMIKHLERCEHRIMESLAWLSDSPFLDLIKOSKDRGPTDHLSESACPL- 461
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 92 FGRNVQERSAGATANPLRSGR---NMEVSLVRRVPLNLPORFGRTTTAKSVCRMLSDLC 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 462 -----ATSAF-----QTKPLKSTLSLFYKKV 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 149 QGSMHSPCANDLFYSMTCHOEIQNPQKQSRRLFFKKI 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 -----ATSAF-----QTKPLKSTLSLFYKKV 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-469-522-4
; Sequence 4, Application US/09469522
; Patent No. US20020151461A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong-Ji
;             Hu, Shi-Xue
;             Benedict, William F.
;             Zhou, Yunli
; TITLE OF INVENTION: MODIFIED RETINOBLASTOMA TUMOR SUPPRESSOR
; PROTEINS
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/469,522
; FILING DATE: 22-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/026,459
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: UTXC:506
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-469-522-4

Query Match          7.1%; Score 72; DB 10; Length 816;
Best Local Similarity 21.5%; Pred. No. 4;
Matches 47; Conservative 23; Mismatches 63; Indels 86; Gaps 8;

```

```

Best Local Similarity 21.5%; Pred. No. 32;
Matches 47; Conservative 23; Mismatches 63; Indels 86; Gaps 8;

OY 17 SSLTSTNIF-----CADELVMS-----NLHSHENYDKYSEPR 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 SKLNDNFHMSLLACALEVVMATYRSTSONLDSGTDLSFPFWILNVLNKAQDFYKVE 421
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 49 GYPKGRSLNFEELK-----DWGPKNVKMTSPVANKPHSPANLPLR 91
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 SFKAEGNLTREMIKHLERCEHRIMESLAWLSDSPFLDLIKOSKDRGPTDHLSESACPL- 480
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 92 FGRNVQERSAGATANPLRSGR---NMEVSLVRRVPLNLPORFGRTTTAKSVCRMLSDLC 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 -----NLPLNNHTAADMYLSPVRS---PKKGSTTRVNSTANAETQ-- 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 149 QGSMHSPCANDLFYSMTCHOEIQNPQKQSRRLFFKKI 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 -----ATSAF-----QTKPLKSTLSLFYKKV 542
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-09-469-522-35
; Sequence 35, Application US/09469522
; Patent No. US20020151461A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong-Ji
;             Hu, Shi-Xue
;             Benedict, William F.
;             Zhou, Yunli
; TITLE OF INVENTION: MODIFIED RETINOBLASTOMA TUMOR SUPPRESSOR
; PROTEINS
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/469,522
; FILING DATE: 22-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/026,459
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hibler, David W.
; REGISTRATION NUMBER: 41,071
; REFERENCE/DOCKET NUMBER: UTXC:506
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-469-522-35

Query Match          7.1%; Score 72; DB 10; Length 832;
Best Local Similarity 21.5%; Pred. No. 33;
Matches 47; Conservative 23; Mismatches 63; Indels 86; Gaps 8;

OY 17 SSLTSTNIF-----CADELVMS-----NLHSHENYDKYSEPR 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Thu Mar 27 11:08:44 2003

Db 378 SKLLNDNFHMSLLACALEVVNATYSRSTSONLDSGTDLSFPWILNVNLKAFDFYKVIE 437  
Qy 49 GYPKGRSLNFEELK-----DWGPKNVIKMSTPAYNKKPHSPANLPLR 91  
Db 438 SFIRAEGLTREMILKHLERCEHRIMESLAWLSDSPFLDLIKOSKDREGPTDHLSESACPL- 496  
Qy 92 FGRNVQERSACATANLPLRSGR---NMEVSLVRVPNLPORFGRTTTAKSVCRMLSDLC 148  
Db 497 -----NLPLQNNHTAADMYLSPVRS---PKKGSTTRVNSTANAETQ-- 535  
Qy 149 QGSMHSPCANDLFYSMTCOHOEIQNDOKQSRRLLEKKI 187  
Db 536 -----ATSAF-----QTKPLKSTLSLEYKKV 558

Search completed: March 26, 2003, 10:23:51  
Job time : 20 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 10:08:29 ; Search time 35 seconds  
(without alignments)  
746.203 Million cell updates/sec

Title: US-09-831-758A-8  
Perfect score: 1018  
Sequence: 1 MEISSLKILLTATSSLL.....KQSRLLFKKIDDAELKQEK 196

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1018	100.0	196	21	Novel human G-prot
2	1018	100.0	196	22	Amino acid sequenc
3	1018	100.0	196	23	Human retina speci
4	1014	99.6	196	23	Human retina speci
5	940	92.3	180	21	Novel human G-prot
6	940	92.3	180	22	Amino acid sequenc
7	674	66.2	196	21	Novel bovine G-pro.
8	674	66.2	196	22	Amino acid sequenc
9	552.5	54.3	203	21	Novel rat G-protei
10	552.5	54.3	203	22	Amino acid sequenc

11	546.5	53.7	203	21	AAV93143	Novel rat G-protei
12	546.5	53.7	203	22	AAG67750	Amino acid sequenc
13	517	50.8	188	21	AAV93144	Novel mouse G-prot
14	517	50.8	188	22	AAG67751	Amino acid sequenc
15	98	9.6	354	21	AAV93144	Arabidopsis thalia
16	98	9.6	426	21	AAV93144	Arabidopsis thalia
17	98	9.6	426	23	AAV93144	Herbicidally activ
18	98	9.6	426	23	AAV93144	Arabidopsis thalia
19	88	8.6	1429	22	ABV58779	Drosophila melanog
20	85	8.3	1436	22	ABV58779	Drosophila melanog
21	83.5	8.2	452	22	AAV93144	Human protein sequ
22	83.5	8.2	1817	21	AAV93144	Human OREF ORF1994
23	82.5	8.1	408	23	ABP27526	Streptococcus poly
24	82.5	8.1	4019	22	AAE13839	Human lung tumour-
25	81.5	8.0	1435	22	ABV11808	Human BC19 homolog
26	81	8.0	1362	22	AAU38416	Salmonella typhi c
27	78	7.7	753	22	ABV60959	Drosophila melanog
28	77	7.6	775	23	AAU93161	Arabidopsis trans
29	76.5	7.5	530	17	AAV92715	Mouse muscle-local
30	76.5	7.5	860	17	AAV92716	Mouse muscle-local
31	76.5	7.5	868	17	AAV92717	Mouse muscle-local
32	76	7.5	296	21	AAV92717	Arabidopsis thalia
33	76	7.5	313	21	AAV92717	Arabidopsis thalia
34	75.5	7.4	371	21	AAV92717	Arabidopsis thalia
35	75.5	7.4	552	22	ABV1902	Drosophila melanog
36	75.5	7.4	869	18	AAV26611	Human muscle-speci
37	75.5	7.4	869	18	AAV26611	Human Dmk receptor
38	75.5	7.4	869	22	AAV77856	Protein of muscle
39	75.5	7.4	869	22	AAV68421	Amino acid sequenc
40	75	7.4	296	21	AAV52208	Arabidopsis thalia
41	75	7.4	313	21	AAV52207	Arabidopsis thalia
42	75	7.4	517	23	ABV53402	Lactococcus lactis
43	75	7.4	725	23	AAE14506	Human mitotic cent
44	75	7.4	928	16	AAV71680	Retinoblastoma ppl
45	75	7.4	928	16	AAV71681	Recombinant ppl10R

## ALIGNMENTS

RESULT 1  
AAV93141  
ID AAV93141 standard; Protein; 196 AA.  
AC AAV93141;  
DT 06-DEC-2000 (first entry)  
DE Novel human G-protein coupled receptor #2.  
KW G-protein coupled receptor; human; bovine; nervous system disorder;  
KW rat; mouse; somatostatin excretion.  
OS Homo sapiens.  
PN WO200029441-A1.  
PD 25-MAY-2000.  
PF 11-NOV-1999; 99WO-JP06283.  
PR 13-NOV-1998; 98JP-0323759.  
PR 08-MAR-1999; 99JP-0060030.  
PR 14-APR-1999; 99JP-0106812.  
PR 14-JUN-1999; 99JP-0166672.  
PR 04-AUG-1999; 99JP-0221640.  
PR 14-SEP-1999; 99JP-0259818.  
(TAKE ) TAKEDA CHEM IND LTD.  
XX Watanabe T, Kikuchi K, Terao Y, Shintani Y, Hinuma S, Fukusumi S;  
XX Fujii R, Hosoya M, Kitada C;

DR WPI; 2000-387747/33.  
 DR N-PSDB; AAA70501.  
 XX G protein coupled-receptor protein and antibodies to it for treatment  
 PT and diagnosis of nerve diseases  
 XX  
 PS Claim 2; Fig 3; 184pp; Japanese.  
 XX  
 CC The invention relates to the isolation of novel G-protein coupled  
 CC receptor (GPCR) genes and their encoded proteins. This sequence  
 CC represents the protein sequence of a human GPCR. The DNAs and proteins  
 CC are used for the treatment, prevention and diagnosis of disorders of  
 CC the nervous system. The proteins and its fragments are also promoters  
 CC of somatostatin excretion.  
 XX  
 SQ Sequence 196 AA;  
 Query Match 100.0%; Score 1018; DB 21; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 5e-106;  
 Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MEIISKFLILLTATSSLLTSNIFCADELVMSNLHSENKYDSEPRGYPKGRSLNFE 60  
 Db 1 MEIISKFLILLTATSSLLTSNIFCADELVMSNLHSENKYDSEPRGYPKGRSLNFE 60  
 Qy 61 ELKDWGPNVKIMSTPAVNKMPHSFANLPLRFRNVQERSAGATANLPLRSGRNMEVSL 120  
 Db 61 ELKDWGPNVKIMSTPAVNKMPHSFANLPLRFRNVQERSAGATANLPLRSGRNMEVSL 120  
 Qy 121 VRRVNLPRQFGRTTAKSVCRMLSDLCQSGSMHSPCANDLFYSMTQCHOEIQNDPKQSR 180  
 Db 121 VRRVNLPRQFGRTTAKSVCRMLSDLCQSGSMHSPCANDLFYSMTQCHOEIQNDPKQSR 180  
 Qy 181 RLLFKKIDDAELKQEK 196  
 Db 181 RLLFKKIDDAELKQEK 196  
 RESULT 2  
 AAG67748  
 ID AAG67748 standard; Protein; 196 AA.  
 AC AAG67748;  
 XX  
 DT 10-DEC-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human prolactin secretion regulating agent.  
 XX Human; prolactin secretion; hypooovarianism; sperm development;  
 KW osteoporosis; lactation disorder; hypothyroidism; kidney failure;  
 KW menopause; hyperprolactinemia; pituitary tumour; dienecephalon tumour;  
 KW menstrual disorder; stress; autoimmune disease; prolactinoma; sterility;  
 KW impotence; amenorrhea; lactorrhea; hyperpituitarism; Sheehan's syndrome  
 KW Chairi-Frommel syndrome; Argonz-del Castillo syndrome; lymphoma;  
 KW Forbes-Albright syndrome; spermatogenesis disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200166134-A1.  
 PN  
 XX  
 XX 13-SEP-2001.  
 PD  
 XX  
 PF 06-MAR-2001; 2001WO-JP01716.  
 XX  
 PF 06-MAR-2000; 2000JP-0065752.  
 PR  
 PR 07-DEC-2000; 2000JP-0378001.  
 XX  
 XX (TAKE ) TAKEDA CHEM IND LTD.  
 PA  
 XX Matsumoto Y, Watanabe T, Hinuma S, Habata Y, Yoshida H;  
 PI  
 XX WPI: 2001-596812/67.  
 DR  
 DR N-PSDB; AAH78481.

XX New polypeptide, useful as a vasotropic, tranquilizing,  
 PT immunosuppressive and gynecological agent comprises the prolactin  
 PT secretion regulator  
 XX  
 PS Claim 2; Fig 3; 180pp; Japanese.  
 XX  
 CC The present sequence represents a human polypeptide which is a prolactin  
 CC secretion regulating agent. The prolactin secretion regulating agent  
 CC polypeptide and polynucleotide are used for the treatment and prevention  
 CC of hypooovarianism, sperm under development, osteoporosis, menopause,  
 CC lactation disorders, hypothyroidism, kidney failure, hyperprolactinemia,  
 CC pituitary tumour, dienecephalon tumour, menstrual disorders, stress,  
 CC autoimmune disease, prolactinoma, sterility, impotence, amenorrhea,  
 CC lactorrhea, hyperpituitarism, Chairi-Frommel syndrome, Argonz-del  
 CC Castillo syndrome, Forbes-Albright syndrome, lymphoma, Sheehan's syndrome  
 CC and spermatogenesis disorders.  
 XX  
 SQ Sequence 196 AA;  
 Query Match 100.0%; Score 1018; DB 22; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 5e-106;  
 Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MEIISKFLILLTATSSLLTSNIFCADELVMSNLHSENKYDSEPRGYPKGRSLNFE 60  
 Db 1 MEIISKFLILLTATSSLLTSNIFCADELVMSNLHSENKYDSEPRGYPKGRSLNFE 60  
 Qy 61 ELKDWGPNVKIMSTPAVNKMPHSFANLPLRFRNVQERSAGATANLPLRSGRNMEVSL 120  
 Db 61 ELKDWGPNVKIMSTPAVNKMPHSFANLPLRFRNVQERSAGATANLPLRSGRNMEVSL 120  
 Qy 121 VRRVNLPRQFGRTTAKSVCRMLSDLCQSGSMHSPCANDLFYSMTQCHOEIQNDPKQSR 180  
 Db 121 VRRVNLPRQFGRTTAKSVCRMLSDLCQSGSMHSPCANDLFYSMTQCHOEIQNDPKQSR 180  
 Qy 181 RLLFKKIDDAELKQEK 196  
 Db 181 RLLFKKIDDAELKQEK 196  
 RESULT 3  
 AAU99162  
 ID AAU99162 standard; Protein; 196 AA.  
 XX  
 AC AAU99162;  
 XX  
 DT 24-SEP-2002 (first entry)  
 XX  
 DE Human retina specific protein C7orf9 variant protein.  
 XX Human; MPP4; C7orf9; C12orf7; F379; retina specific gene;  
 KW AMD; age-related macular degeneration; blindness; gene therapy;  
 KW Ophthalmological; transgenic.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 32  
 FT /note= "Wild-type Ile substituted by Met"  
 XX  
 XX WO200244366-A2.  
 PN  
 XX  
 XX 06-JUN-2002.  
 PD  
 XX  
 PF 29-NOV-2001; 2001WO-EP13940.  
 XX  
 PF 29-NOV-2000; 2000US-253751P.  
 PR  
 XX (MULT-) MULTIGENE BIOTECH GMBH.  
 PA  
 XX Stoehr HB, Weber BHF;  
 PI  
 XX

DR WPI: 2002-508512/54.  
 XX Novel nucleic acids encoding retina-specific human protein C7orf9,  
 PT C12orf7, MPP4 or F379, useful for diagnosing age-related macular  
 PT degeneration or predisposition for macular degeneration, and in gene  
 PT therapy techniques -  
 XX Example 2; Page -: 120pp; English.  
 XX The invention relates to isolated nucleic acid encoding retina-specific  
 CC human protein C7orf9, C12orf7, MPP4 or F379 or a fragment, derivative or  
 CC allelic variation of the above mentioned nucleic acid sequences.  
 CC Also included are a recombinant vector containing the nucleic acid, a  
 CC recombinant host cell which contains the vector and expresses the  
 CC protein, an inhibitor characterised in that it can suppress the  
 CC activity of the protein, treating macular degeneration or a  
 CC predisposition for macular degeneration, comprising administering to a  
 CC mammalian subject a reagent which decreases, inhibits or increases  
 CC expression of C7orf9, C12orf7, MPP4 and/or F379 or which leads to the  
 CC expression of a biologically active C7orf9, C12orf7, MPP4 and/or F379  
 CC protein and a transgenic non-human animal comprising at least one of the  
 CC nucleic acids (active or inactivated). The nucleic acid or protein is  
 CC useful for diagnosing macular degeneration, preferably age-related  
 CC macular degeneration (AMD) or a predisposition for macular degeneration.  
 CC The reagent used in the diagnosis is a C7orf9-, C12orf7-, MPP4- or F379  
 CC -specific nucleic acid probe, or anti-C7orf9, anti-C12orf7, anti-MPP4 or  
 CC anti-F379-antibody. The reagent is detectably labeled, with a compound  
 CC such as a radioisotope, a bioluminescent compound, a chemiluminescent  
 CC compound, a fluorescent compound, a metal chelate or an enzyme. Fragments  
 CC of the nucleic acid are useful as probes or primers in a diagnostic  
 CC assay, and for identifying further factors involved in development and  
 CC progression of macular degeneration. The proteins encoded by the nucleic  
 CC acid are useful to identify further unrelated proteins which are  
 CC associated with macular degeneration and for use in screening methods  
 CC based on protein/protein interactions. The nucleic acid is also useful as  
 CC reagents for detecting differences between normal and aberrant expression  
 CC of the protein. The nucleic acid is also useful in gene therapy  
 CC techniques, and can be used for gene targeting and/or gene replacement  
 CC for restoring a mutant gene or for creating a mutant gene via homologous  
 CC recombination. The protein can be used to identify other proteins  
 CC involved in development and progression of macular degeneration.  
 CC The present sequence represents a variant of the retina specific protein  
 CC C7orf9.  
 CC Note: The present sequence is not shown in the specification but  
 CC was created by the indexer using the sequence appearing as ABG66919 and  
 CC the information in example 2.  
 XX Sequence 196 AA;  
 SQ  
 Query Match 100.0%; Score 1018; DB 23; Length 196;  
 Best Local Similarity 100.0%; Pred. No. 5e-106;  
 Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEITSSKFLILLTATSSLTNIFCADELVMNLSKENVKDYSEPRGYPKGRSLNFE 60  
 Db 1 MEITSSKFLILLTATSSLTNIFCADELVMNLSKENVKDYSEPRGYPKGRSLNFE 60  
 QY 61 ELKDWGPKNVTKMSTPAVNKMPHSFANLPLRFGNRVQERSAGATANLPLRSGRNMEVSL 120  
 Db 61 ELKDWGPKNVTKMSTPAVNKMPHSFANLPLRFGNRVQERSAGATANLPLRSGRNMEVSL 120  
 QY 121 VRRVPLNLPQRFGRITAKSVCRMLSDLCQSMHSPCANLFLYSMTQCHQETQNDQKQSR 180  
 Db 121 VRRVPLNLPQRFGRITAKSVCRMLSDLCQSMHSPCANLFLYSMTQCHQETQNDQKQSR 180  
 QY 181 RLLEFKKIDDAELKQEK 196  
 Db 181 RLLEFKKIDDAELKQEK 196

RESULT 4  
 ABG66919  
 ID ABG66919 standard; Protein: 196 AA.

XX ABG66919;  
 XX 24-SEP-2002 (first entry)  
 XX Human retina specific protein encoded by cDNA C7orf9.  
 XX Human; MPP4; C7orf9; C12orf7; F379; retina specific gene;  
 KW AMD; age-related macular degeneration; blindness; gene therapy;  
 KW Ophthalmological; transgenic.  
 XX Homo sapiens.  
 XX WO200244366-A2.  
 XX 06-JUN-2002.  
 XX 29-NOV-2001; 2001WO-EP13940.  
 XX 29-NOV-2000; 2000US-253751P.  
 XX (MULT-) MULTIGENE BIOTECH GMBH.  
 XX Stoehr HB, Weber BHF;  
 WPI: 2002-508512/54.  
 N-PSDB; ABK95399.  
 XX Novel nucleic acids encoding retina-specific human protein C7orf9,  
 PT C12orf7, MPP4 or F379, useful for diagnosing age-related macular  
 PT degeneration or predisposition for macular degeneration, and in gene  
 PT therapy techniques -  
 XX Claim 1; Fig 9; 120pp; English.  
 XX The invention relates to isolated nucleic acid encoding retina-specific  
 CC human protein C7orf9, C12orf7, MPP4 or F379 or a fragment, derivative or  
 CC allelic variation of the above mentioned nucleic acid sequences.  
 CC Also included are a recombinant vector containing the nucleic acid, a  
 CC recombinant host cell which contains the vector and expresses the  
 CC protein, an inhibitor characterised in that it can suppress the  
 CC activity of the protein, treating macular degeneration or a  
 CC predisposition for macular degeneration, comprising administering to a  
 CC mammalian subject a reagent which decreases, inhibits or increases  
 CC expression of C7orf9, C12orf7, MPP4 and/or F379 or which leads to the  
 CC expression of a biologically active C7orf9, C12orf7, MPP4 and/or F379  
 CC protein and a transgenic non-human animal comprising at least one of the  
 CC nucleic acids (active or inactivated). The nucleic acid or protein is  
 CC useful for diagnosing macular degeneration, preferably age-related  
 CC macular degeneration (AMD) or a predisposition for macular degeneration.  
 CC The reagent used in the diagnosis is a C7orf9-, C12orf7-, MPP4- or F379  
 CC -specific nucleic acid probe, or anti-C7orf9, anti-C12orf7, anti-MPP4 or  
 CC anti-F379-antibody. The reagent is detectably labeled, with a compound  
 CC such as a radioisotope, a bioluminescent compound, a chemiluminescent  
 CC compound, a fluorescent compound, a metal chelate or an enzyme. Fragments  
 CC of the nucleic acid are useful as probes or primers in a diagnostic  
 CC assay, and for identifying further factors involved in development and  
 CC progression of macular degeneration. The proteins encoded by the nucleic  
 CC acid are useful to identify further unrelated proteins which are  
 CC associated with macular degeneration and for use in screening methods  
 CC based on protein/protein interactions. The nucleic acid is also useful as  
 CC reagents for detecting differences between normal and aberrant expression  
 CC of the protein. The nucleic acid is also useful in gene therapy  
 CC techniques, and can be used for gene targeting and/or gene replacement  
 CC for restoring a mutant gene or for creating a mutant gene via homologous  
 CC recombination. The protein can be used to identify other proteins  
 CC involved in development and progression of macular degeneration.  
 CC The present sequence represents a retina specific protein of  
 CC the invention.  
 XX Sequence 196 AA;  
 SQ  
 Query Match 99.6%; Score 1014; DB 23; Length 196;

Best Local Similarity 99.5%; Pred. No. 1.4e-105;  
Matches 195; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEIISKFLILLTATSSLLTSNIFCADELVMSNLHSEKDYKSEPRGYPKGERSLNFE 60  
DB 1 MEIISKFLILLTATSSLLTSNIFCADELVMSNLHSEKDYKSEPRGYPKGERSLNFE 60  
QY 61 ELKDMGPKNVIKMSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNNEVSL 120  
DB 61 ELKDMGPKNVIKMSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNNEVSL 120  
QY 121 VRRVNLPRFQRTTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTCHQHOEIQNPOKQSR 180  
DB 121 VRRVNLPRFQRTTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTCHQHOEIQNPOKQSR 180  
QY 181 RLFFKKIDDAELKQEK 196  
DB 181 RLFFKKIDDAELKQEK 196

RESULT 5  
AAY93140  
ID AAY93140 standard; Protein; 180 AA.  
XX  
AC AAY93140;  
XX  
DT 06-DEC-2000 (first entry)  
XX  
DE Novel human G-protein coupled receptor #1.  
XX  
KW G-protein coupled receptor; human; bovine; nervous system disorder;  
KW rat; mouse; somatostatin excretion.  
XX  
OS Homo sapiens.  
XX  
PN WO200029441-A1..  
XX  
PD 25-MAY-2000.  
XX  
PF 11-NOV-1999; 99WO-JP06283.  
XX  
PR 13-NOV-1998; 98JP-0323759.  
PR 08-MAR-1999; 99JP-0060030.  
PR 14-APR-1999; 99JP-0106812.  
PR 14-JUN-1999; 99JP-0166672.  
PR 04-AUG-1999; 99JP-0221640.  
PR 14-SEP-1999; 99JP-0259818.  
XX  
PA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
PI Watanabe T, Kikuchi K, Terao Y, Shintani Y, Hinuma S, Fukusumi S;  
PI Fujii R, Hosoya M, Kitada C;  
XX  
DR WPI: 2000-387747/33.  
DR N-PSDB; AAY70500.  
XX  
G protein coupled receptor protein and antibodies to it for treatment and diagnosis of nerve diseases  
XX  
PS Claim 1; Fig 1; 184pp; Japanese.  
XX  
CC The invention relates to the isolation of novel G-protein coupled receptor (GPCR) genes and their encoded proteins. This sequence represents the protein sequence of a human GPCR. The DNAs and proteins are used for the treatment, prevention and diagnosis of disorders of the nervous system. The proteins and its fragments are also promoters of somatostatin excretion.  
XX  
SQ Sequence 180 AA;

Query Match 92.3%; Score 940; DB 21; Length 180;  
Best Local Similarity 100.0%; Pred. No. 2.6e-97;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEIISKFLILLTATSSLLTSNIFCADELVMSNLHSEKDYKSEPRGYPKGERSLNFE 60  
DB 1 MEIISKFLILLTATSSLLTSNIFCADELVMSNLHSEKDYKSEPRGYPKGERSLNFE 60  
QY 61 ELKDMGPKNVIKMSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNNEVSL 120  
DB 61 ELKDMGPKNVIKMSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNNEVSL 120  
QY 121 VRRVNLPRFQRTTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTCHQHOEIQNPOKQSR 180  
DB 121 VRRVNLPRFQRTTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTCHQHOEIQNPOKQSR 180

RESULT 6  
AAG67747  
ID AAG67747 standard; Protein; 180 AA.  
XX  
AC AAG67747;  
XX  
DT 10-DEC-2001 (first entry)  
XX  
DE Amino acid sequence of a human prolactin secretion regulating agent.  
XX  
KW Human; prolactin secretion; hypovarianism; sperm development;  
KW osteoporosis; lactation disorder; hypothyroidism; kidney failure;  
KW hyperprolactinemia; pituitary tumour; dienecephalon tumour; menopause;  
KW menstrual disorder; stress; autoimmune disease; prolactinoma; sterility;  
KW impotence; amenorrhea; lactorrhea; hyperpituitarism; Sheehan's syndrome  
KW Chari-Frommel syndrome; Argonz-del Castillo syndrome; lymphoma;  
KW Forbes-Albright syndrome; spermatogenesis disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200166134-A1.  
XX  
PD 13-SEP-2001.  
XX  
PF 06-MAR-2001; 2001WO-JP01716.  
XX  
PR 06-MAR-2000; 2000JP-0065752.  
PR 07-DEC-2000; 2000JP-0378001.  
XX  
PA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
PI Matsumoto Y, Watanabe T, Hinuma S, Habata Y, Yoshida H;  
XX  
DR WPI: 2001-596812/67.  
DR N-PSDB; AAY78475.  
XX  
PT New polypeptide, useful as a vasotropic, tranquillizing,  
PT immunosuppressive and gynecological agent comprises the prolactin  
PT secretion regulator  
XX  
PS Claim 1; Fig 1; 180pp; Japanese.  
XX  
CC The present sequence represents a human polypeptide which is a prolactin secretion regulating agent. The prolactin secretion regulating agent polypeptide and polynucleotide are used for the treatment and prevention of hypovarianism, sperm under development, osteoporosis, menopause, lactation disorders, hypothyroidism, kidney failure, hyperprolactinemia, pituitary tumour, dienecephalon tumour, menstrual disorders, stress, autoimmune disease, prolactinoma, sterility, impotence, amenorrhea, lactorrhea, hyperpituitarism, Chari-Frommel syndrome, Argonz-del Castillo syndrome, Forbes-Albright syndrome, lymphoma, Sheehan's syndrome and spermatogenesis disorders.  
XX  
SQ Sequence 180 AA;

Query Match 92.3%; Score 940; DB 22; Length 180;  
Best Local Similarity 100.0%; Pred. No. 2.6e-97;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEIISSKLFIILLTLATSSLLTSNIFCADELVMNLHSHKENVKYSRGPYKPGKRSNLEE 60  
Db 1 MEIISSKLFIILLTLATSSLLTSNIFCADELVMNLHSHKENVKYSRGPYKPGKRSNLEE 60  
QY 61 ELKDWGPKNVKIMSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNMEVSL 120  
Db 61 ELKDWGPKNVKIMSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNMEVSL 120  
QY 121 VRRVNPILPQRFGRTRTTAKSVCRMLSDLCQSGMHSPCANDLFYSMTQHQEIQNDQKOSR 180  
Db 121 VRRVNPILPQRFGRTRTTAKSVCRMLSDLCQSGMHSPCANDLFYSMTQHQEIQNDQKOSR 180

RESULT 7  
AAY93142  
ID AAY93142 standard; Protein: 196 AA.  
XX AC AAY93142;  
XX DT 06-DEC-2000 (first entry)  
XX DE Novel bovine G-protein coupled receptor.  
XX KW G-protein coupled receptor; human; bovine; nervous system disorder;  
XX KW rat; mouse; somatostatin excretion.  
XX OS Bos taurus.  
XX PN WO200029441-A1.  
XX PD 25-MAY-2000.  
XX PF 11-NOV-1999; 99WO-JP06283.  
XX PR 13-NOV-1998; 98JP-0323759.  
XX PR 08-MAR-1999; 99JP-0060030.  
XX PR 14-APR-1999; 99JP-0106812.  
XX PR 14-JUN-1999; 99JP-0166672.  
XX PR 04-AUG-1999; 99JP-0221640.  
XX PR 14-SEP-1999; 99JP-0259818.  
XX PA (TAKE ) TAKEDA CHEM IND LTD.  
XX PI Watanabe T, Kikuchi K, Terao Y, Shintani Y, Hinuma S, Fukusumi S;  
XX PI Fujii R, Hosoya M, Kitada C;  
XX DR WPI; 2000-387747/33.  
XX DR N-PSDB; AAA70502.  
XX PT G protein coupled receptor protein and antibodies to it for treatment  
XX PT and diagnosis of nerve diseases  
XX PS Claim 2; Fig 4; 184pp; Japanese.  
XX CC The invention relates to the isolation of novel G-protein coupled  
XX CC receptor (GPCR) genes and their encoded proteins. This sequence  
XX CC represents the protein sequence of a bovine GPCR. The DNAs and proteins  
XX CC are used for the treatment, prevention and diagnosis of disorders of  
XX CC the nervous system. The proteins and its fragments are also promoters  
XX CC of somatostatin excretion.  
XX SX Sequence 196 AA;  
Query Match 66.2%; Score 674; DB 21; Length 196;  
Best Local Similarity 71.9%; Pred. No. 2.4e-67;  
Matches 143; Conservative 13; Mismatches 37; Indels 6; Caps 3;

QY 1 MEIISSKLFIILLTLATSSLLTSNIFCADELVMNLHSHKENVKYSRGPYKPGKRSNLEE 57  
Db 1 MEIISSKLFIILLTLATSSLLTSNIFCADELVMNLHSHKENVKYSRGPYKPGKRSNLEE 57  
QY 58 NFEELKDWGPKNVKIMSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNME 117  
Db 58 NFEELKDWGPKNVKIMSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNME 117

Db 60 TFEVVDWAPK--IKMKNPVYKMPSPSAANLPLRFRGNMEEERSTRAMAHLPRLQKRE 117  
QY 118 VSLVRRVNPILPQRFGRTRTTAKSVCRMLSDLCQSGMHSPCANDLFYSMTQHQEIQNDQK 177  
Db 118 VSLVRRVNPILPQRFGRTRTTAKSVCRMLSDLCQSGMHSPSTNGLLYSMACQPEIQNDQK 177  
QY 178 QSRRLFLFKKIDDAELKOEK 196  
Db 178 NLRRRGFOKIDDAELKOEK 196

RESULT 8  
AAG67749  
ID AAG67749 standard; Protein: 196 AA.  
XX AC AAG67749;  
XX DT 10-DEC-2001 (first entry)  
XX DE Amino acid sequence of a bovine prolactin secretion regulating agent.  
XX KW Bovine prolactin secretion; hypovarianism; sperm development;  
XX KW osteoporosis; lactation disorder; hypothyroidism; kidney failure;  
XX KW menopause; hyperprolactinemia; pituitary tumour; dienecephalon tumour;  
XX KW menstrual disorder; stress; autoimmune disease; prolactinoma; sterility;  
XX KW impotence; amenorrhea; lactorrhea; hyperpituitarism; Sheehan's syndrome  
XX KW Chairi-Frommel syndrome; Argonz-del Castillo syndrome; lymphoma;  
XX KW Forbes-Albright syndrome; spermatogenesis disorder.  
XX OS Bos sp.  
XX PN WO200166134-A1.  
XX PD 13-SEP-2001.  
XX PF 06-MAR-2001; 2001WO-JP01716.  
XX PR 06-MAR-2000; 2000JP-0065752.  
XX PR 07-DEC-2000; 2000JP-0378001.  
XX PA (TAKE ) TAKEDA CHEM IND LTD.  
XX PI Matsumoto Y, Watanabe T, Hinuma S, Habata Y, Yoshida H;  
XX PI WPI; 2001-596812/67.  
XX DR N-PSDB; AAH78486.  
XX PT New polypeptide, useful as a vasotropic, tranquilizing,  
XX PT immunosuppressive and gynecological agent comprises the prolactin  
XX PT secretion regulator  
XX PS Claim 2; Fig 4; 180pp; Japanese.  
XX CC The present sequence represents a bovine polypeptide which is a prolactin  
XX CC secretion regulating agent. The prolactin secretion regulating agent  
XX CC polypeptide and polynucleotide are used for the treatment and prevention  
XX CC of hypovarianism, sperm under development, osteoporosis, menopause,  
XX CC lactation disorders, hypothyroidism, kidney failure, hyperprolactinemia,  
XX CC pituitary tumour, dienecephalon tumour, menstrual disorders, stress,  
XX CC autoimmune disease, prolactinoma, sterility, impotence, amenorrhea,  
XX CC lactorrhea, hyperpituitarism, Chairi-Frommel syndrome, Argonz-del  
XX CC Castillo syndrome, Forbes-Albright syndrome, lymphoma, Sheehan's syndrome  
XX CC and spermatogenesis disorders.  
XX SX Sequence 196 AA;  
Query Match 66.2%; Score 674; DB 22; Length 196;  
Best Local Similarity 71.9%; Pred. No. 2.4e-67;  
Matches 143; Conservative 13; Mismatches 37; Indels 6; Caps 3;

QY 1 MEIISSKLFIILLTLATSSLLTSNIFCADELVMNLHSHKENVKYSRGPYKPGKRSNLEE 57  
Db 1 MEIISSKLFIILLTLATSSLLTSNIFCADELVMNLHSHKENVKYSRGPYKPGKRSNLEE 57

QY 58 NPEELKDMGPKNVKMKSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNME 117  
 DB 60 TEEVYKDWAPK--IKMKPVVKNMPPSAANLPLRGRNMEERSTRAMHLPRLGKNRE 117  
 QY 118 VSLRRVNLPRFORGRTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTQCHOEIQNDQK 177  
 DB 118 DLSLRWPNLPRFORGRTTAKSITKLSNLLQSQSMHSPSTNGLLYSMACOPQOEIQNDQK 177  
 QY 178 QSRLLFKKIDDAELKQEK 196  
 DB 178 NLRRGFKIDDAELKQEK 196

RESULT 9  
 ID AAY93145 standard; Protein; 203 AA.

AC AAY93145;  
 DT 06-DEC-2000 (first entry)  
 DE Novel rat G-protein coupled receptor #2.  
 XX G-protein coupled receptor; human; bovine; nervous system disorder;  
 KW rat; mouse; somatostatin excretion.  
 XX Rattus sp.  
 OS WO200029441-A1.  
 PN 25-MAY-2000.  
 XX 11-NOV-1999; 99WO-JP06283.  
 XX 13-NOV-1998; 98JP-0323759.  
 PR 08-MAR-1999; 98JP-0060030.  
 PR 14-APR-1999; 99JP-0108812.  
 PR 14-JUN-1999; 99JP-0166672.  
 PR 04-AUG-1999; 99JP-0221640.  
 PR 14-SEP-1999; 99JP-0259818.  
 XX (TAKE ) TAKEDA CHEM IND LTD.  
 PA Watanabe T, Kikuchi K, Terao Y, Shintani Y, Hinuma S, Fukusumi S;  
 PI Fujii R, Hosoya M, Kitada C;  
 XX WPI: 2000-387747/33.  
 DR N-PSDB; AAA70505.

PT G protein coupled receptor protein and antibodies to it for treatment  
 PT and diagnosis of nerve diseases  
 XX Claim 2; Page 176-177; 184pp; Japanese.  
 CC The invention relates to the isolation of novel G-protein coupled  
 CC receptor (GPCR) genes and their encoded proteins. This sequence  
 CC represents the protein sequence of a rat GPCR. The DNAs and proteins  
 CC are used for the treatment, prevention and diagnosis of disorders of  
 CC the nervous system. The proteins and its fragments are also promoters  
 CC of somatostatin excretion.

XX Sequence 203 AA;  
 SQ Query Match 54.3%; Score 552.5; DB 21; Length 203;  
 Best Local Similarity 58.1%; Pred. No. 1.2e-53;  
 Matches 115; Conservative 26; Mismatches 46; Indels 11; Gaps 3;

QY 1 MEIISKLFIILLATSSLLTSNIFCADELVMNLHNSKENYDKYSPGYPKG--ERSLN 58  
 DB 1 MEIISKRFIILLATSSLLTSNIFCADELVMNLHNSKENYDKYSPGYPKG--ERSLN 60  
 QY 59 FEELKDMGPKNVKMKSTPAVNKMPHSFANLPLRFRGNVQERSAGATANLPLRSGRNMEV 118

DB 61 FOELKDMGAKKDIKMSAPANKVPHSAANLPLRFRGNIEDRRSPARA-----NMEA 112  
 QY 119 SLVRRVNLPRFORGRTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTQCHOEIQNDQK 178  
 DB 113 GTMHHFSLPRFORG--TTARRITKTLAGLPKSLHSLASSELLYAMTRHQEIQSGOEO 171  
 QY 179 SRRLLFKKIDDAELKQEK 196  
 DB 172 PKRVFETDDAERKQEK 189

RESULT 10  
 ID AAG67756 standard; Protein; 203 AA.

AC AAG67756;  
 DT 10-DEC-2001 (first entry)  
 DE Amino acid sequence of a rat prolactin secretion regulating agent.  
 XX Rat; prolactin secretion; hypovarianism; sperm development;  
 KW osteoporosis; lactation disorder; hypothyroidism; kidney failure;  
 KW menopause; hyperprolactinemia; pituitary tumour; dienecephalon tumour;  
 KW menstrual disorder; stress; autoimmune disease; prolactinoma; sterility;  
 KW impotence; amenorrhea; lactorrhea; hyperpituitarism; Sheehan's syndrome  
 KW Chairi-Frommel syndrome; Argonz-del Castillo syndrome; lymphoma;  
 KW Forbes-Albright syndrome; spermatogenesis disorder.  
 XX Rattus sp.  
 OS WO200166134-A1.  
 PN 13-SEP-2001.  
 XX 06-MAR-2001; 2001WO-JP01716.  
 PR 06-MAR-2000; 2000JP-0065752.  
 PR 07-DEC-2000; 2000JP-0378001.  
 XX (TAKE ) TAKEDA CHEM IND LTD.  
 PA Matsumoto Y, Watanabe T, Hinuma S, Habata Y, Yoshida H;  
 PI WPI: 2001-596812/67.  
 DR N-PSDB; AAH78515.

PT New polypeptide, useful as a vasotropic, tranquilizing,  
 PT immunosuppressive and gynecological agent comprises the prolactin  
 PT secretion regulator  
 XX Claim 2; Page 166-167; 180pp; Japanese.

XX The present sequence represents a rat polypeptide which is a prolactin  
 CC secretion regulating agent. The prolactin secretion regulating agent  
 CC polypeptide and polynucleotide are used for the treatment and prevention  
 CC of hypovarianism, sperm under development, osteoporosis, menopause,  
 CC lactation disorders, hypothyroidism, kidney failure, hyperprolactinemia,  
 CC pituitary tumour, dienecephalon tumour, menstrual disorders, stress,  
 CC autoimmune disease, prolactinoma, sterility, impotence, amenorrhea,  
 CC lactorrhea, hyperpituitarism, Chairi-Frommel syndrome, Argonz-del  
 CC Castillo syndrome, Forbes-Albright syndrome, lymphoma, Sheehan's syndrome  
 CC and spermatogenesis disorders.

XX Sequence 203 AA;  
 SQ Query Match 54.3%; Score 552.5; DB 22; Length 203;  
 Best Local Similarity 58.1%; Pred. No. 1.2e-53;  
 Matches 115; Conservative 26; Mismatches 46; Indels 11; Gaps 3;

QY 1 MEIISKLFIILLATSSLLTSNIFCADELVMNLHNSKENYDKYSPGYPKG--ERSLN 58  
 DB 1 MEIISKRFIILLATSSLLTSNIFCADELVMNLHNSKENYDKYSPGYPKG--ERSLN 59  
 QY 118 DLSLRWPNLPRFORGRTTAKSITKLSNLLQSQSMHSPSTNGLLYSMACOPQOEIQNDQK 177

QY	59	FEELDKWGPKNVIKMTSTPAVNKMPHSFANLPLRFGRNVQBERSAGATANLPLRSGRNMEV	111
DB	61	FOELDKWGAKDKTKMSPAPANKVPHSANLPLRFGRNIEDRSPARA	112
QY	119	SLVRRYPNLPRFGRTTTAKSYCRMSLSDLCQSMHSPCANDLFYSMTQHQBIQNPDKQ	178
DB	113	GTMSHEFSLPQRFGR-TTARRITKTLAGLPQKSLHSLASSESLYAMTRQHQEIQSPGOEQ	171
QY	179	SRLLEPKKIDDAELKQEK	196
DB	172	PKRKFVETETDDAERKQEK	189
RESULT 12			
QY	AG67750		
ID	AAAG67750 standard; Protein; 203 AA.		
XX	AAAG67750:		
AC	XX		
CC	XX		
DT	XX		
DE	10-DEC-2001 (first entry)		
XX	Amino acid sequence of a rat prolactin secretion regulating agent.		
XX	Bovine; prolactin secretion; hypovarianism; sperm development;		
XX	osteoporosis; lactation disorder; hypothyroidism; kidney failure;		
KW	menopause; hyperprolactinemia; pituitary tumour; diencephalon tumour;		
KW	menstrual disorder; stress; autoimmune disease; prolactinoma; sterili		
KW	impotence; amenorrhea; lactorrhea; hyperpituitarism; Sheehan's syndro		
KW	Chairi-Frommel syndrome; Argonz-del Castillo syndrome; lymphoma;		
KW	Forbes-Albright syndrome; spermatogenesis disorder.		
OS	Rattus sp.		
XX	WO200166134-A1.		
XX	13-SEP-2001.		
XX	06-MAR-2001; 2001WO-JP01716.		
XX	06-MAR-2000; 2000JP-0065752.		
FR	07-DEC-2000; 2000JP-0378001.		
XX	(TAKE ) TAKEDA CHEM IND LTD.		
XX	Matsumoto Y, Watanabe T, Hinuma S, Habata Y, Yoshida H:		
PI	WPI; 2001-596812/67.		
DR	N-PSDB; AAH78489.		
XX	New polypeptide, useful as a vasotropic, tranquilizing,		
PT	immunosuppressive and gynecological agent comprises the prolactin		
PT	secretion regulator		
XX	Claim 2; Fig 5; 180pp; Japanese.		
PS	The present sequence represents a rat polypeptide which is a prolactin		
XX	secretion regulating agent. The prolactin secretion regulating agent		
CC	polypeptide and polynucleotide are used for the treatment and prevent		
CC	of hypovarianism, sperm under development, osteoporosis, menopause,		
CC	lactation disorders, hypothyroidism, kidney failure, hyperprolactinemia,		
CC	pituitary tumour, diencephalon tumour, menstrual disorders, stress,		
CC	autoimmune disease, prolactinoma, sterility, impotence, amenorrhea,		
CC	lactorrhea, hyperpituitarism, Chairi-Frommel syndrome, Argonz-del		
CC	Castillo syndrome, Forbes-Albright syndrome, lymphoma, Sheehan's syndr		
CC	and spermatogenesis disorders.		
XX	Sequence 203 AA;		
SQ	Query Match 53.7%; Score 546.5; DB 22; Length 203;		
	Best Local Similarity 57.6%; Pred. No. 5.7e-53;		
	Matches 114; Conservative 26; Mismatches 47; Indels 11; Gaps		

Db 1 MEIISRRFILLTATSSFLTSNTLCDELMPHFHSKREGYKYYQLRGIPKGVKERSVT 60  
 QY 59 FEELKDWGPKNVKIMSTPAVNKMPHSFANPLRFRGNVQERSAGATANPLRSGRNNEV 118  
 Db 61 FOELKDWGAKKIDIKMSPAPANKVPKHSANLPLRFGRTTIDKRSAPARY-----NMEA 112  
 QY 119 SLVRVPNLPORFGRTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTCOHOEIONPOKQ 178  
 Db 113 GTSMHFFSLPQRFGR-TTARRITKTLAGLPKSLHSLASSESLYAMTROHOEIQSPGQEQ 171  
 QY 179 SRLLFKKIDDAELKQEK 196  
 Db 172 PKRVFTETDDAERKQEK 189

RESULT 13  
 AAY93144  
 ID AAY93144 standard; Protein; 188 AA.  
 AC AAY93144;  
 XX  
 DT 06-DEC-2000 (first entry)  
 DE  
 DE Novel mouse G-protein coupled receptor #1.  
 KW G-protein coupled receptor; human; bovine; nervous system disorder;  
 KW rat; mouse; somatostatin excretion.  
 OS Mus sp.  
 XX WO200029441-A1.  
 XX 25-MAY-2000.  
 XX 11-NOV-1999; 99WO-JP06283.  
 XX 13-NOV-1998; 98JP-0323759.  
 XX 08-MAR-1999; 99JP-0060030.  
 XX 14-APR-1999; 99JP-0106812.  
 XX 14-JUN-1999; 99JP-0166672.  
 XX 04-AUG-1999; 99JP-0221640.  
 XX 14-SEP-1999; 99JP-0259818.  
 XX (TAKE ) TAKEDA CHEM IND LTD.  
 XX Watanabe T, Kikuchi K, Terao Y, Shintani Y, Hinuma S, Fukusumi S;  
 PI Fujii R, Hosoya M, Kitada C;  
 XX WPI; 2000-387747/33.  
 XX N-PSDB; AAA70504.  
 XX G protein coupled receptor protein and antibodies to it for treatment  
 PT and diagnosis of nerve diseases -  
 XX  
 XX Claim 2; Fig 7; 184pp; Japanese.  
 XX  
 CC The invention relates to the isolation of novel G-protein coupled  
 CC receptor (GPCR) genes and their encoded proteins. This sequence  
 CC represents the protein sequence of a mouse GPCR. The DNAs and proteins  
 CC are used for the treatment, prevention and diagnosis of disorders of  
 CC the nervous system. The proteins and its fragments are also promoters  
 CC of somatostatin excretion.  
 XX  
 XX Sequence 188 AA;

Query Match 50.8%; Score 517; DB 21; Length 188;  
 Best Local Similarity 57.1%; Pred. No. 1e-49; Length 188;  
 Matches 113; Conservative 22; Mismatches 51; Indels 12; Gaps 4;  
 QY 1 MEIISRRFILLTATSSFLTSNTLCDELMPHFHSKREGYKYYQLRGIPKGVKERSVT 60  
 Db 1 MEIISRRFILLTATSSFLTSNTLCDELMPHFHSKREGYKYYQLRGIPKGVKERSVT 60

QY 59 FEELKDWGPKNVKIMSTPAVNKMPHSFANPLRFRGNVQERSAGATANPLRSGRNNEV 118  
 Db 61 FOELKDWGAKKIDIKMSPAPANKVPKHSANLPLRFGRTTIDKRSAPARY-----NMEA 112  
 QY 119 SLVRVPNLPORFGRTTAKSVCRMLSDLCOGSMHSPCANDLFYSMTCOHOEIONPOKQ 178  
 Db 113 GTSMHFFSLPQRFGR-TTARRITKTLAGLPKSLHSLASSESLYAMTROHOEIQSPGQEK 170  
 QY 179 SRLLFKKIDDAELKQEK 196  
 Db 171 TRGAFTETDDAERKPEK 188

RESULT 14  
 AAG67751  
 ID AAG67751 standard; Protein; 188 AA.  
 AC AAG67751;  
 XX  
 DT 10-DEC-2001 (first entry)  
 DE  
 DE Amino acid sequence of a murine prolactin secretion regulating agent.  
 KW Human; prolactin secretion; hypovarianism; sperm development;  
 KW osteoporosis; lactation disorder; hypothyroidism; kidney failure;  
 KW menopause; hyperprolactinemia; pituitary tumour; dienecephalon tumour;  
 KW menstrual disorder; stress; autoimmune disease; prolactinoma; sterility;  
 KW impotence; amenorrhea; lactorrhea; hyperpituitarism; Sheehan's syndrome  
 KW Chauri-Frommel syndrome; Argonz-del Castillo syndrome; lymphoma;  
 KW Forbes-Albright syndrome; spermatogenesis disorder.  
 XX Mus sp.  
 XX WO200166134-A1.  
 XX 13-SEP-2001.  
 XX 06-MAR-2001; 2001WO-JP01716.  
 XX 06-MAR-2000; 2000JP-0065752.  
 XX 07-DEC-2000; 2000JP-0378001.  
 XX (TAKE ) TAKEDA CHEM IND LTD.  
 XX Matsumoto Y, Watanabe T, Hinuma S, Habata Y, Yoshida H;  
 PI WPI; 2001-596812/67.  
 XX N-PSDB; AAH78503.  
 XX New polypeptide, useful as a vasotropic, tranquilizing,  
 PT immunosuppressive and gynecological agent comprises the prolactin  
 PT secretion regulator -  
 XX  
 XX Claim 2; Page 156-157; 180pp; Japanese.

XX The present sequence represents a murine polypeptide which is a prolactin  
 CC secretion regulating agent. The prolactin secretion regulating agent  
 CC polypeptide and polynucleotide are used for the treatment and prevention  
 CC of hypovarianism, sperm under development, osteoporosis, menopause,  
 CC lactation disorders, hypothyroidism, kidney failure, hyperprolactinemia,  
 CC pituitary tumour, dienecephalon tumour, menstrual disorders, stress,  
 CC autoimmune disease, prolactinoma, sterility, impotence, amenorrhea,  
 CC lactorrhea, hyperpituitarism, Chauri-Frommel syndrome, Argonz-del  
 CC Castillo syndrome, Forbes-Albright syndrome, lymphoma, Sheehan's syndrome  
 CC and spermatogenesis disorders.  
 XX  
 XX Sequence 188 AA;

Query Match 50.8%; Score 517; DB 22; Length 188;  
 Best Local Similarity 57.1%; Pred. No. 1e-49;  
 Matches 113; Conservative 22; Mismatches 51; Indels 12; Gaps 4;

QY 1 MEIISKLILLTATSSLLTSNTFCADLVMSNLHSHKENDKYSEPRGPKG--ERSLN 58  
 Db 1 MEIISLARFILLVATSSFTNTCTDEPMFHSKEDGKYSQLRGIPKGEKERSVS 60  
 QY 59 FEELKDWGPKNVKIMKSPAYNKMPHSPANLPLRFRGNVQEERSAGATANPLRSGRNMEV 118  
 Db 61 FOELKDWGAKNVKIMSPAPANKVPHSAANLPLRFRGTIDEKRSPPAARV-----NNEA 112  
 QY 119 SLVRRVNLQRFQRTTAKSVCRMLSDLCQGSWHSPCANDLFLYSNTCQHQEIQNPQKQ 178  
 Db 113 GTRSHFSLQRFQR--TTARSP-KTPADLPQKPLHSLGSSLELYVMTCQHQEIQSPGGKR 170  
 QY 179 SRRLFKKIDDAELKOEK 196  
 Db 171 TRGAFVETDDAERKPEK 188

RESULT 15  
 AAG44544  
 ID AAG44544 standard; Protein; 354 AA.  
 XX AAG44544;  
 AC  
 DT 18-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 55810.  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 hybridisation assay; genetic mapping; gene expression control; promoter;  
 termination sequence.  
 KW Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127482.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 21-JUN-1999; 99US-0139763.  
 PR 22-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 22-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147036.  
 PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 16-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 9.6%; Score 98; DB 21; Length 354;  
Best Local Similarity 29.3%; Pred. No. 0.036;  
Matches 53; Conservative 20; Mismatches 66; Indels 42; Gaps 11;  
QY 2 EIISSKLFILLTATSSLLTSNIFCADELVMNSLNHSENKYSEP---RCYPKGERSLN 58  
Db 13 OLISAKEDIKVLRTK-----FCHPILVRLGHWHDAGTYNNKNEEPLRGANG--SLR 63  
QY 59 FE-ELK---DWGPKNVIKMSTPAYNKMPH-SFANLPLRFRGNVOERSAGATANLPLRSG 113  
Db 64 FEELKHAANAGLLNALKLIQPLKDKYPNISIYADL-----FOLASATA-IEEAGG 112  
QY 114 RNMEVSLVRRVNPQLPQRFGR-TTAKSVCRMLSDLCQSMHSPC--ANDLFYSMTQHOE 170  
Db 113 -----PDIPMKYGRVDVVAPEQCPEEGRLPDAGPPSPADHLRDFYRMGLDDKE 161  
QY 171 I 171  
Db 162 I 162

Search completed: March 26, 2003, 10:13:55  
Job time : 36 secs